(12)

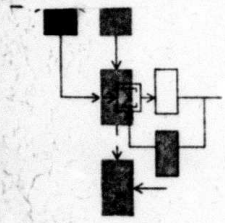
NOVEMBER 1983

LIDS-TH-1341

Research Supported By:

Defense Advanced Projects Agency
Grant ONR-N00014-75-C-1183

National Science Foundation
Grant NSF/ECS-79-19880



Fundamental Issues of Multiple Accessing

Joseph Yu Ngai Nui

Laboratory for Information and Decision Systems
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS 02139

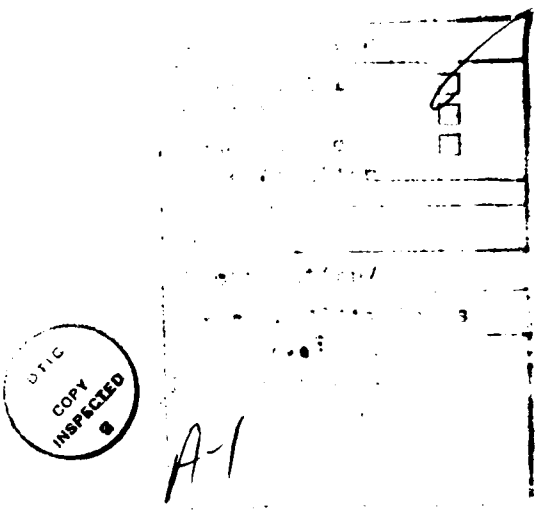
November 1983 LIDS-TH-1341

FUNDAMENTAL ISSUES OF MULTIPLE ACCESSING

by

Joseph Yu Ngai Hui

This report is based on the unaltered thesis of Joseph Yu Ngai Hui, submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, November 1983. The research was conducted at the M.I.T. Laboratory for Information and Decision Systems, with support provided in part by the Defense Advanced Projects Agency under Grant ONR-N00014-75-C-1183 and by the National Science Foundation under Grant NSF/ECS-79-19880.



Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. A136276 / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) FUNDAMENTAL ISSUES OF MULTIPLE ACCESSING	5. TYPE OF REPORT & PERIOD COVERED Thesis
	6. PERFORMING ORG. REPORT NUMBER LIDS-TH-1341
7. AUTHOR(s) Joseph Yu Ngai Hui	8. CONTRACT OR GRANT NUMBER(s) ARPA Order No. 3045/5-7-75 ONR/N00014-75-C-1183
9. PERFORMING ORGANIZATION NAME AND ADDRESS Massachusetts Institute of Technology Laboratory for Information and Decisions Systems Cambridge, Massachusetts 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Code No. 5T10 ONR Identifying No. 049-383
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209	12. REPORT DATE November 1983 / 13. NUMBER OF PAGES 230
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research Information Systems Program Code 437 Arlington, Virginia 22217	15. SECURITY CLASS. (of this report) UNCLASSIFIED / 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Fundamental issues of multiple accessing are identified. These issues include transmitter asynchronism, variability of the set of active users, feedback, and degree of codebook knowledge among the users. Various multiple access schemes are examined under these issues. These issues are subsequently modeled and analysed using an information theoretic framework.

We discover that the capacity region of the asynchronous multiple acces channel is different from that of the synchronous channel.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. (Continued).

For communication systems with users having random message generation time, we demonstrate that its channel capacity resembles that of an asynchronous channel even though the users are synchronous, if decoding delay is constrained to be much smaller than the message inter-arrival time.

We investigage communication with restricted decoder structure. New information theoretic quantities that incorporate the decoding metric used are discovered and examined. Using these quantities, we provide a rigorous and novel treatment for the theory of jamming. These mathematical techniques provide insight for achieving reliable communication in a multiple access environment where each user may not know the codebook of the other users and a jammer may be present.

We then apply the general theory developed to three specific asynchronous channel without feedback, namely the OR channel, the spread spectrum channel and the collision channel. Practical and novel coding schemes are suggested. Maximum throughput, error rate and decoding complexity are analysed.

FUNDAMENTAL ISSUES OF MULTIPLE ACCESSING

Joseph Yu Ngai Hui

S.B., Massachusetts Institute of Technology
(1981)

S.M., Massachusetts Institute of Technology
(1981)

E.E., Massachusetts Institute of Technology
(1982)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

October 1983

Signature of Author.....Joseph Yu Ngai Hui.....
Department of Electrical Engineering
and Computer Science

Certified by.....Pierre A. Humblet.....
Pierre A. Humblet, Thesis Supervisor

Accepted by.....
Chairman, Departmental Committee on Graduate Students

FUNDAMENTAL ISSUES OF MULTIPLE ACCESSING

Joseph Yu Ngai Hui

Submitted to the
Department of Electrical Engineering and Computer Science
on October 6, 1983 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy

ABSTRACT

Fundamental issues of multiple accessing are identified. These issues include transmitter asynchronism, variability of the set of active users, feedback, and degree of codebook knowledge among the users. Various multiple access schemes are examined under these issues. These issues are subsequently modeled and analysed using an information theoretic framework.

We discover that the capacity region of the asynchonous multiple access channel is different from that of the synchronous channel.

For communication systems with users having random message generation time, we demonstrate that its channel capacity resembles that of an asynchronous channel even though the users are synchronous, if decoding delay is constrained to be much smaller than the message inter-arrival time.

We investigate communication with restricted decoder structure. New information theoretic quantities that incorporate the decoding metric used are discovered and examined. Using these quantities, we provide a rigorous and novel treatment for the theory of jamming. These mathematical techniques provide insight for achieving reliable communication in a multiple access environment where each user may not know the codebook of the other users and a jammer may be present.

We then apply the general theory developed to three specific asynchronous channel without feedback, namely the OR channel, the spread spectrum channel and the collision channel. Practical and novel coding schemes are suggested. Maximum throughput, error rate and decoding complexity are analysed.

Thesis supervisor: Dr. Pierre A. Humblet

Title: NEC associate professor of computers and communication.

Acknowledgement

Many people contribute to my formal education, equipping me with a body of technical knowledge, the methods of scientific inquiry, the language of scientific expression, and a professional outlook. Thus I would like to thank here all those who help to instill these attributes in my professional career.

I thank my thesis supervisor, Professor Pierre Humblet for his critical evaluation of ideas, which helps to produce more definitive and convincing versions of my doctoral dissertation progressively over the past two years, and more generally, sharpens my research and expressive skills. His suggestions bring tremendous improvements to the final version of this thesis.

I also thank my thesis readers, Professors Robert Gallager and Peter Elias, for their insights and comments. Professor Gallager is instrumental to my graduate career and in arranging my future employment with the Bell Laboratories, Murray Hill, N.J.

Professors Wilbur Davenport, Jeff Shapiro, George Verghese (my graduate academic advisor), Herman Haus (my undergraduate academic advisor) are particularly helpful, both academically and personally, in various stages of my stay at MIT.

My education and my research contributions are impossible without financial support over the years. For this, I am

particularly grateful to MIT for providing generous financial aids for my undergraduate studies. I would also like to acknowledge the funding of this research by the National Science Foundation (NSF/ECS 79-19880) and the Advanced Research Project Agency (ONR/N00014-75-C1183).

The Laboratory for Information and Decision Systems, where this research is performed, provides friendly and excellent technical and administrative support. My fellow graduate students in the Laboratory also make graduate studies memorable and pleasant.

Farewell MIT. Farewell Boston. Farewell my acquaintances here. Thanks for a truly educational and enlightening episode in my life.

My parents deserve much of the credit for my doctorate. Their conviction to equip their children with the more noble goods of skill and character continues to motivate us. Such conviction is testified by their material and emotional sacrifices for us. My long absence from home is one such example. To them I dedicate my thesis.

TABLE OF CONTENTS

Chapter 1 Issues of Multiple Accessing

1.1 Multiple accessing schemes

Multiple accessing is the technique of communication by a group of users over a shared channel. An information theoretic model of the multiple access channel was given by Shannon [1]. Such a model, however, has failed to yield useful results for the various multiple access schemes used in practice. Such schemes include Time Division Multiple Accessing (TDMA), the Aloha scheme [2], the Capetanakis tree algorithm [20] (modified by Gallager [3]) and Code Division Multiple Accessing (CDMA)[4,5,6]. This variety can be attributed to different requirements and resources of the communication system. The purpose of this thesis is to clarify the fundamental issues of multiple accessing so that a more comprehensive information theoretic model can be developed. Coding theorems are stated and proved for these issues. Practical and novel multiple access schemes for specific channels are suggested and analysed.

It would be illuminating to look at specific multiple access schemes before we abstract the issues involved in multiple accessing. First of all, we shall introduce some notations. Let M be the number of users in the system. Each user transmits symbols from the alphabet X. The channel output alphabet is Y. We assume that the symbols transmitted by the M users are synchronous (a condition to be relaxed later for individual channels). The channel output y, given that the M users transmit

the symbols $x_1, x_2, \ldots, x_M$ respectively, occurs according to the probabilities $P(y/x_1 x_2 \ldots x_M)$. The throughput of a scheme is the sum of the information rates for all M users, normalized by the maximum throughput for the case when the channel is used by a single user. Using these notations, the four schemes mentioned in the previous paragraph and the classical multiple access channel of Shannon can be characterized as follows

Scheme 1 Time division multiple accessing

Channel: The collision channel defined as follows.
X = {2^n packets each of length n} U {idle}
Y = {2^n packets each of length n} U {idle} U {collision}
$P(y/x_1 \ldots x_M)$: $y = x_i$ if for some i, x_i = packet and x_j = idle for all $j \neq i$; y = idle if all x_i = idle; y = collision if more than one x_i's are packets.

M: fixed

Access scheme: Time is divided into frames. Each frame consists of M slots. Each user transmits its packet in its assigned slot. The user transmits idle if it does not transmit a packet.

Timing: Slot synchronization and frame synchronization required.

Throughput: 1

Scheme 2 The slotted Aloha scheme [2]

Channel: The collision channel defined previously.

M: Large, each user generates packets with Poisson statistics.

Access scheme: Transmit a packet once generated. If the transmission results in a collision, the user retransmits the packet after a random delay. Repeated retransmission until the packet is successfully transmitted.

Timing: Slot synchronization required for slotted Aloha.

Throughput: $e^{-1} = 0.368$. This simple scheme, however, is unstable in the sense that the channel soon becomes flooded with retransmission.

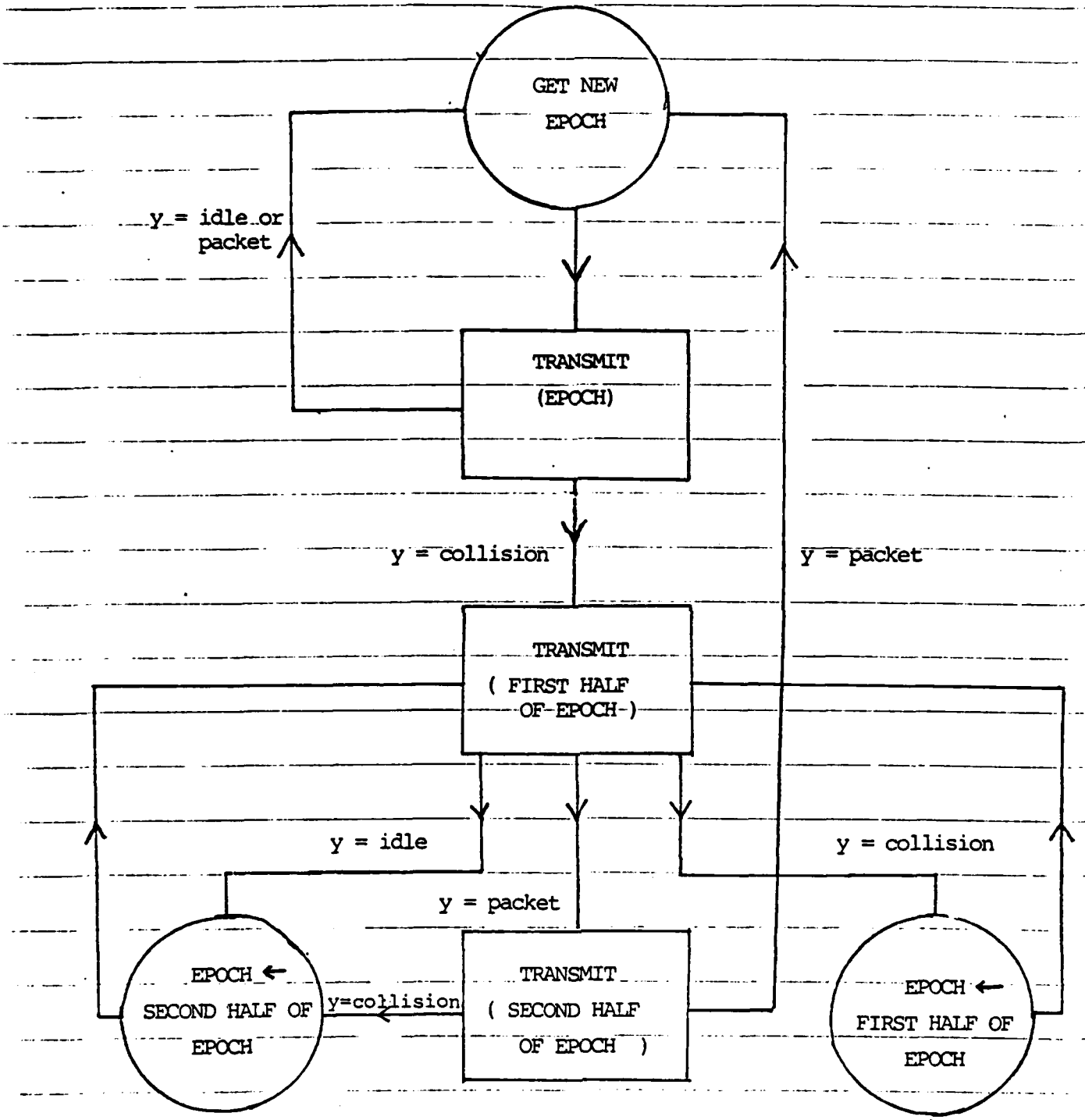
Effect of symbol asynchronism: When the packets are not synchronized, we have the pure Aloha scheme. A packet is totally erased if it collides, even partially, with other packets. The throughput of pure Aloha is $e^{-1}/2$.

Scheme 3 The Tree algorithm (improved version of Gallager [3])

Channel: The collision channel

M: Large, each user generates packets with Poisson statistics with total rate λ for all users.

Access scheme: each packet is labeled with its generation time. The algorithm tries to resolve conflicts resulting from transmitting packets in a time period called an epoch. The algorithm works as shown in figure 1.1.1. There are two kinds of



GET NEW EPOCH: Epoch start from end of last epoch with length 1.08λ.

TRANSMIT(*) : All packets with generation time in the period * transmit in the next slot.

Figure 1.1.1 The modified version of the tree algorithm.

states (circle and square) in the flow chart. The circles redefine the epoch. The squares authorize the set of users with generation time in a certain subinterval of the epoch (shown as the argument in the function Transmit(*)) to transmit in the next slot. The channel output of the transmission determines the next state in the flow chart.

Timing: Slot synchronization required.

Throughput: .487

Scheme 4 Code division multiple accessing

We shall consider code division multiple accessing for three types of channel, namely the OR channel, the spread spectrum channel and the collision channel. A more detailed description and analysis of these channels and coding schemes are found in chapters 5, 6 and 7 of this thesis. For this scheme, the number of users can be large, with only a small portion of the users active at a time. Those users that are not active 'transmit' the idle symbol, which we assume to be in X. The message of each user is redundantly coded for achieving reliable communication in the presence of interference due to the signals of the other users. We say that joint decoding is performed if the codewords sent by the other users are also estimated for the sake of obtaining a better estimate of one's message. We say that joint decoding is not performed if the signals of the other users are treated as memoryless noise in the estimation of one's message. No feedback is required for these schemes. Symbol synchronization

(as shall be examined in each case) and frame (a frame is the length of a codeword) synchronization are not required. The three channels are described as follows

A. The OR channel [4]

Channel:

$X = \{0,1\}$, idle = 0

$Y = \{0,1\}$

$P(y/x_1 \ldots x_M)$: $y = 0$ if all $x_i = 0$; 1 otherwise.

Access scheme: One scheme uses pulse position modulation and convolutional encoding. A pulse position modulation symbol is a sequence with a one (called a pulse) among n-1 zeros, and hence there are n pulse positions. The binary (0 or 1) message stream is fed into a convolutional encoder shown in figure 1.1.2 which puts out pulse position modulation symbols. The pulses sent by the other users become interference to a receiver.

Throughput: .69 if joint decoding is not performed; 1 if joint decoding is performed.

The effect of symbol asynchronism: In continuous time, a pulse has a duration of τ , thus symbol asynchronism may occur. By limiting the temporal resolution of the pulse positions and using a decision rule which obtains a discrete estimate of the pulse position, we may model the channel in a symbol synchronous manner.

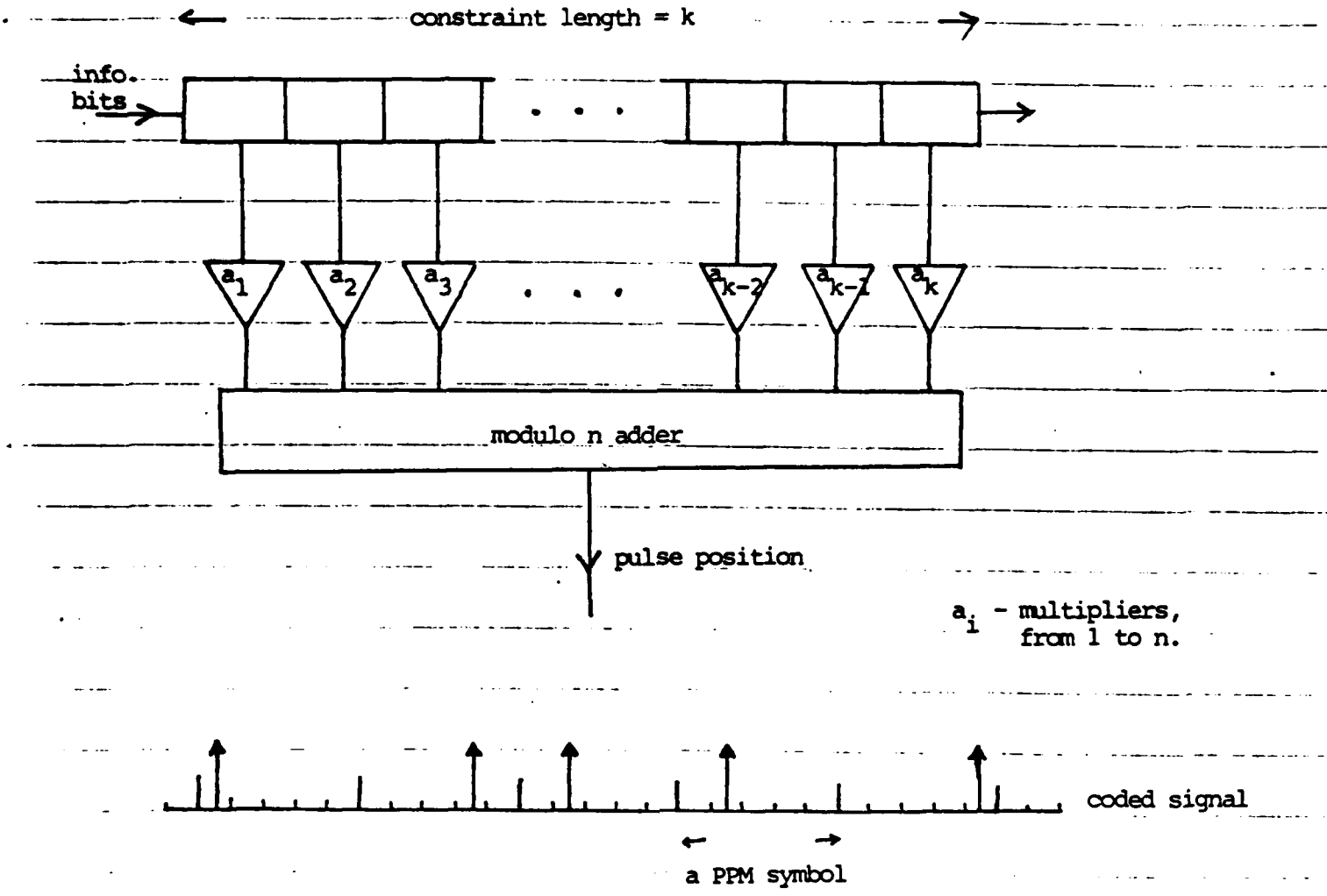


Figure 1.1.2 Coding scheme for the OR channel

B. The spread spectrum channel (or the adder channel) [5]

Channel:

$X = \{-1,+1\} \cup \{0\}$; the idle symbol 0 may not be used during the transmission of a message.

Y = The set of integers

$P(y/x_1 \ldots x_M)$: $y = \sum_{i=1}^{M} x_i$.

Access scheme: Each user is assigned an n bit long Pseudo-Noise (PN) sequence $\{c_{ij}\}_{j=1}^{n}$, $c_{ij} \in \{\pm 1\}$. The binary $\{0,1\}$ message stream $\{\ldots u_{i,-2}\ u_{i,-1}\ u_{i,0}\ u_{i,1}\ u_{i,2}\ \ldots\}$ is encoded as the binary $\{0,1\}$ data stream $\{\ldots x_{i,-2}\ x_{i,-1}\ x_{i,0}\ x_{i,1}\ x_{i,2}\ \ldots\}$. Define $1\{c_{ij}\} = \{c_{ij}\}_{j=1}^{n}$ and $0\{c_{ij}\} = \{-c_{ij}\}_{j=1}^{n}$. The code stream that is sent by the transmitter comprises the sequences $x_{i,k}\{c_{ij}\}$ transmitted successively for the consecutive k's. The coding scheme, realized using continuous waveform, is shown in figure 1.1.3.

Throughput: .721 if joint decoding is not performed.

The effect of symbol asynchronism: In continuous time, a symbol is realized as a chip of duration τ as shown in figure 1.1.3. Symbol asynchronism occurs when the chips are not synchronized for the users. It is shown in chapter 6 that symbol asynchronism does not affect throughput for a specific structure of the demodulator-decoder.

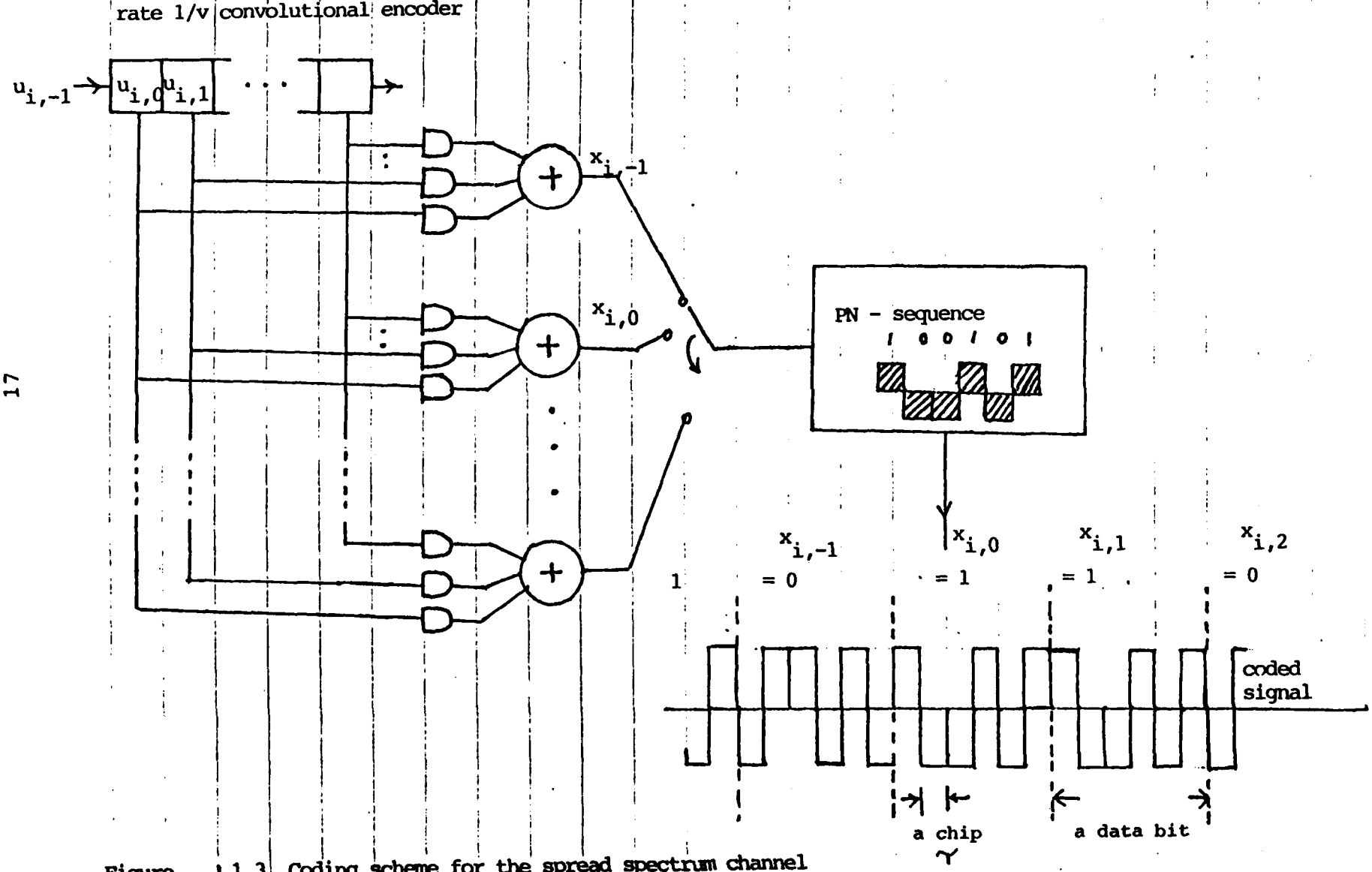


Figure 1.1.3 Coding scheme for the spread spectrum channel

C. The collision channel [6]

Channel: the collision channel

Access scheme: Each user sends packets infrequently and successive packets are encoded. This is achieved by having the message redundantly coded, then interleaved and broken up into packets. Collision results in erasure of the entire packet, but coding enables the reconstruction of the original message.

Throughput: $e^{-1} = .368$

The effect of symbol asynchronism: When the packets are not synchronized, partial overlapping of packets results in total loss of the packets. Using a technique called super-packeting to be described in chapter 7, the e^{-1} throughput is not affected by symbol asynchronism.

The classical multiple access channel [1]

Channel: General

M: fixed

Access scheme: Joint decoding assumed. Coding theorem proved only for the case without feedback. It has been shown that feedback may enlarge the capacity region. (example: The two-user adder channel of Wolf [1] with $X = \{+1,-1\}$ and $y = x_1 + x_2$.)

Timing: Symbol synchronization and frame synchronization required. Recent work by Cover et. al. [7] shows that channel

capacity is not affected if the code frames of the users are shifted slightly with respect to each other.

Throughput: The capacity region for the two-user synchronous multiple access channel is

$$\mathcal{R} = \text{convex hull } \bigcup_{P_1(X_1),P_2(X_2)} \tilde{\mathcal{R}}$$

$$: P(x_1 x_2 y) = P_1(x_1) P_2(x_2) P(y/x_1 x_2)$$

where (R_1, R_2) iff

$$R_1 < I(X_1; Y/X_2)$$

$$R_2 < I(X_2; Y/X_1)$$

$$R_1 + R_2 < I(X_1 X_2; Y)$$

The above result has been generalized for the case of M>2 and the case of correlated sources [1].

Several questions arise if we compare these five schemes closely.

1. The classical theory of multiple accessing gives a throughput of 1 for the collision channel, instead of .368. While the throughput of TDMA agrees with the classical theory, the throughput of the Aloha scheme and the tree algorithm are both less than .5. What accounts for the difference in throughput?

2. What accounts for the higher throughput of the tree algorithm compared with the Aloha algorithm?

3. The use of code division multiple accessing for the collision channel achieves the same throughput as the Aloha scheme, yet without feedback. Can the Aloha channel (that is, the collision channel with feedback about the occurrence of collision) achieve a higher throughput? If not, why is it that the feedback of the Aloha channel cannot increase throughput beyond that of the code division multiple access scheme?

4. Why does feedback enlarge the capacity region of the two-user adder channel of Wolf?

5. How does joint decoding improve the throughput of the OR channel?

We shall next abstract the fundamental issues of multiple accessing and answer the above questions for the purpose of illustrating these issues.

1.2 The issue of synchronization

Two forms of synchronism appear in the examples in the previous section, namely symbol synchronism and frame synchronism. With symbol asynchronism, the characterization of the channel output is quite cumbersome and has to be described individually for each channel. The analysis in this thesis shall assume symbol synchronism except in chapters 5, 6 and 7 when symbol asynchronism is treated individually for the OR channel, the spread spectrum channel and the collision channel.

The more important issue is frame synchronism. Frame synchronism is the degree of agreement of a common time frame among the users. Each user is assumed to have a clock with the same frequency, but may have different initializations. Asynchronism results from the uncertainty about the initializations of the clocks of other users. It is worth emphasizing that we are dealing with asynchronism among the transmitters, not between the transmitter-receiver pairs. Synchronization between the transmitter and the receiver can be achieved by the transmission of preambles for synchronizing the receiver, or by simply synchronizing the clocks at the transmitter and the receiver beforehand.

Why is a common time frame useful? Many multiple access schemes try to reduce conflict or interference among the users so as to improve channel throughput. This sometimes requires a common time frame. For example, TDMA requires perfect frame

synchronization so that the system would be collision free. A common time frame also allows the use of time sharing. Suppose we have two multiple access schemes that achieve two sets of transmission rates for the users. By using the first scheme a fraction of the time and the second scheme the remaining fraction of the time, the convex combination of the two sets of rates can be achieved. This explains the convex hull required for characterizing the capacity region for the classical synchronous multiple access channel. Consider the example of the collision channel. The capacity regions of figure 1.2.1a and b are obviously achievable by allowing a single user to transmit all the time. Figure 1.2.1c gives the convex hull of figure 1.2.1a and b. It happens that this convex hull includes the entire capacity region. Consequently, the rate of one-half is achievable for the two users at the same time.

Synchronization is a matter of degree. A precise definition of the degree of asynchronism will be given in chapter 2 while the intuitive concept is introduced here. Perfect asynchronism happens when each transmitter knows absolutely nothing about the initialization of the clocks of the other transmitters. At the other extreme, we have ideal TDMA for which a common time frame makes perfect elimination of collision possible. In between these two extremes, the clocks may differ by some finite and unknown differences. For the TDMA scheme, we may use guard time to avoid collision. The system will be collision free if the guard time is larger than the maximum possible clock difference. Consequently, a capacity of one can be achieved by

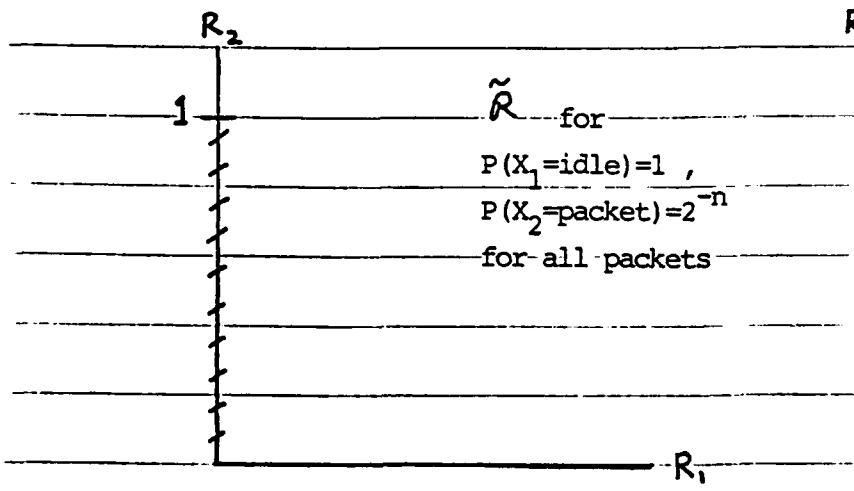


Figure 1.2.1a

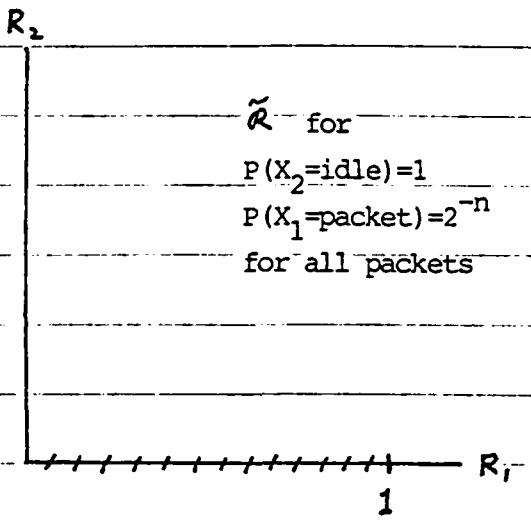


Figure 1.2.1b

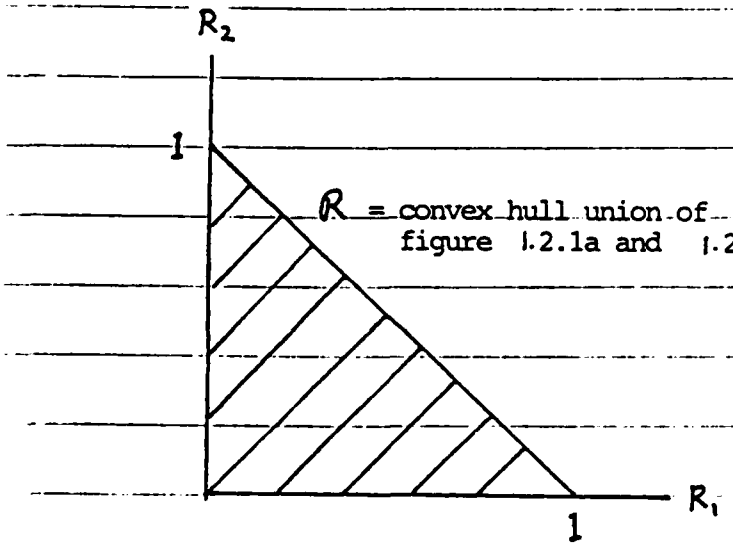


Figure 1.2.1c Capacity region for the synchronous channel

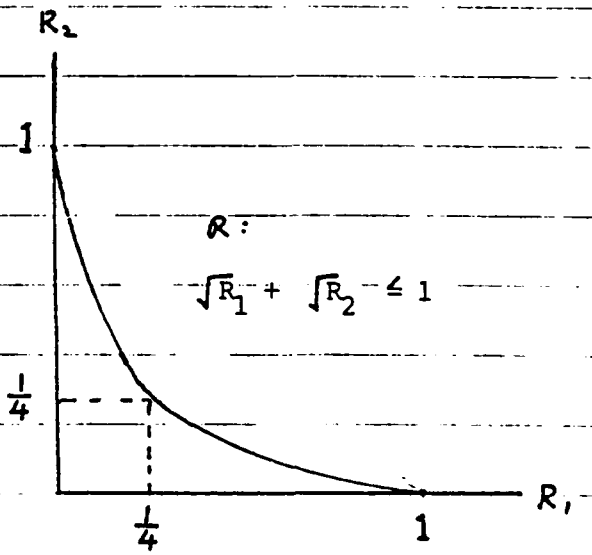


Figure 1.2.1d Capacity region for the asynchronous channel

Figure 1.2.1 Capacity regions for the collision channel

using round-robin TDMA with sufficiently long slots for each user so that the guard time is negligible in proportion. The cost, however, is that the messages suffer a delay (roughly the length of a frame for the case of TDMA) that is much longer than the degree of asynchronism (the minimum guard time required to avoid conflict for TDMA). In general, the capacity region for the multiple access channel with mild asynchronism is the same as that with perfect synchronism. This fact is proved by Cover et. al. [7]. Their proof utilizes time sharing of two codes for the ordinary multiple access channel and uses maximum likelihood decoding over shifts of the hypothesized transmitter codewords as shown in figure 1.2.2. Instead of having a guard time, the two coding frames are allowed to overlap, and the overlapped region is ignored in the decoding. In the proof, the shift d (or equivalently, the length of the overlapped region) is assumed bounded and the length of each frame, say n, is substantially larger than d. Their argument then lets d (and consequently n) go to infinity, with d/n goes to zero.

However, frame synchronization is often inaccurate and difficult to achieve in practice. For users with small data rate, poorly synchronized clocks may require a guard time much longer than the duration of a data burst. To achieve the synchronous capacity would require impractically long frames which cause intolerable delay. With the trend of increasing bandwidth per link and increasing number of users (often transmitting in small bursts), it is often convenient to assume that the channel is perfectly asynchronous. In essence, we want the length of the

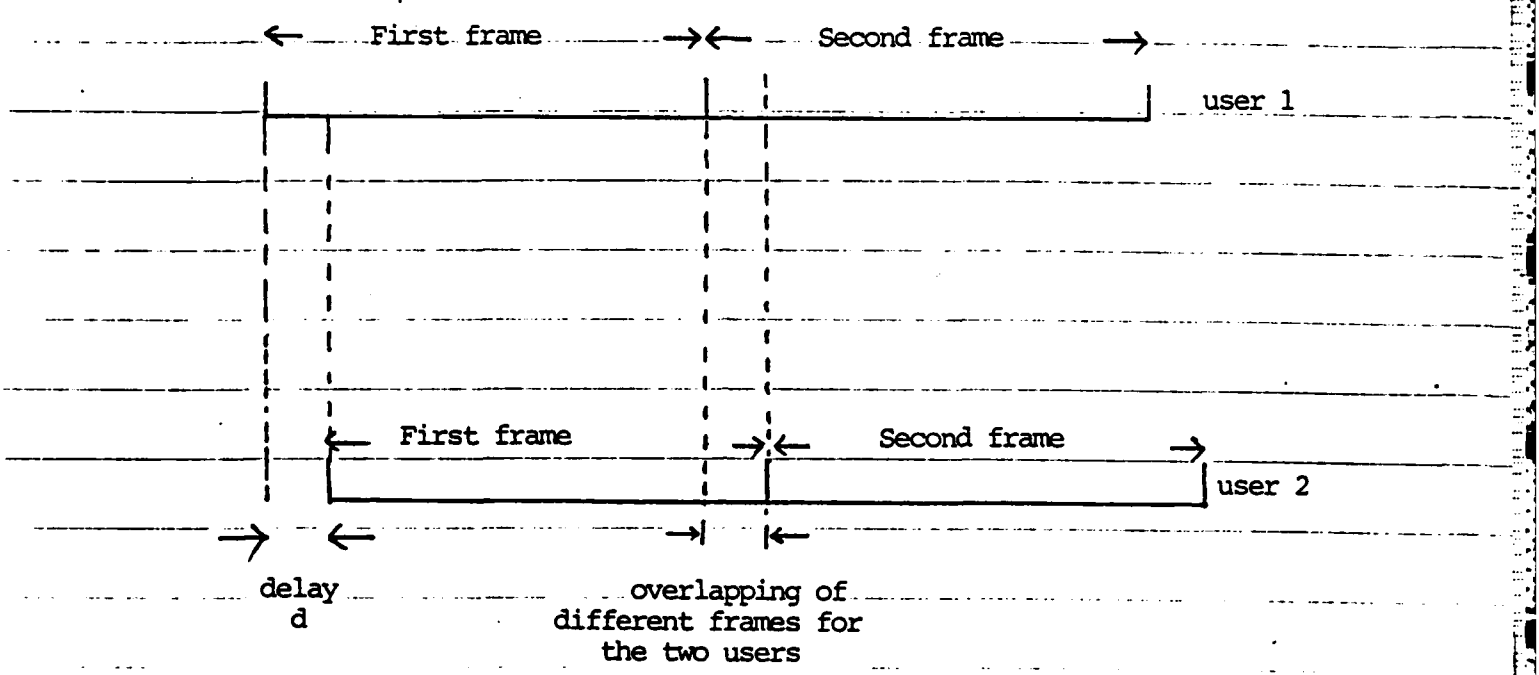


Figure 1.2.2 Coding for the mildly asynchronous channel

codewords n to be smaller than the shifts d.

Chapter 2 proves that the capacity region for the two-user asynchronous multiple access channel without feedback is

$$\mathcal{R} = \bigcup_{P_1(X_1),P_2(X_2) : P(x_1 x_2 y)=P_1(x_1)P_2(x_2)P(y/x_1 x_2)} \tilde{\mathcal{R}}$$

where $(R_1, R_2) \in \tilde{\mathcal{R}}$ iff

$$R_1 < I(X_1;Y/X_2)$$

$$R_2 < I(X_2;Y/X_1)$$

$$R_1 + R_2 < I(X_1 X_2;Y)$$

It is noteworthy that the capacity of the synchronous channel is the convex hull of the capacity region defined above. This fact also holds for more than two users. The time sharing argument, which achieves the convex hull region for the asynchronous channel, cannot be used for the perfectly asynchronous channel.

Consider again the example of the two-user collision channel, this time with perfect asynchronism. The capacity region is given in figure 1.2.1d, which is not convex and is smaller than that of the synchronous channel in figure 1.2.1c. For the M user asynchronous collision channel with equal transmission rates for the users, the total throughput can be shown to be less than $(1-1/M)^{M-1}$, which approaches e^{-1} for large M.

1.3 The issue of uncertainty about the set of simultaneous users

The paradox of e^{-1} versus 1 still remains unexplained. Consider a system with M users, each generating packets at a Poisson rate of λ/M (packets per slot), $0<\lambda<1$. Suppose all M transmitters are perfectly synchronized. With sufficient buffering, a throughput of 1 can be achieved by using round-robin TDMA. No feedback is required. The Aloha scheme and the tree algorithm, both requiring feedback, seem to have a much smaller throughput. Lack of synchronization does not cause such discrepancy because supplying synchronized clocks for each user of the Aloha channel or the tree algorithm does not seem to help. It may be pointed out that M is assumed infinite for these two schemes. However, an infinite M never occurs in practice.

The paradox disappears if the communication system imposes a maximum allowable delay D (in terms of slots) for the reception of the packets from its generation time. More precisely, if D satisfies the condition $M/[2(1-\lambda)] >> D >> 1/\lambda$, then the capacity of the collision channel is e^{-1} even though the users are synchronized. We shall give an intuitive explanation of this result here, while chapter 3 will give a rigorous argument for the general channel.

First, we examine the average delay for round-robin TDMA, which can be viewed as an M server queue with exponential inter-arrival time and deterministic service time. The average

delay can be shown to be $M/2 + M/[2(1-\lambda)] + 1$, in which the first term accounts for the random arrival time of a packet within a round-robin cycle, the second term accounts for the queueing delay, and the third term accounts for the transmission time of the packet over the channel, that is, a slot. Ignoring the transmission time, the average delay simplifies to $M/[2(1-\lambda)]$. For given λ (which may be close to one in order to achieve a throughput close to one), average delay is proportional to M. For systems with large M (say 1000), D becomes intolerably large. We are interested in the case when the maximum tolerable delay D is much smaller than $M/[2(1-\lambda)]$.

On the other hand, we cannot have the maximum tolerable delay too small without decreasing the reliability of communication. The number of packets generated within D equals $\lambda D+O(\sqrt{D\lambda})$. When D is just a few $1/\lambda$'s, the number of packets generated within D is small and highly variable, thus there may be more packets to be transmitted within the allowable delay D than the channel can handle.

Therefore, the case of interest satisfies the condition $M/[2(1-\lambda)] >> D >> 1/\lambda$. Since $M/[2(1-\lambda)] >> D$, most users would have at most one packet to send within D. Since $D >> 1/\lambda$, the set of active users would have size N which is approximately $D\lambda$, by the law of large numbers. Thus, there is an uncertainty about which combination of N users is active within a set of M users.

The obvious question then is how to communicate reliably in the absence of feedback. Conflict free scheduling is

impossible in this case. The solution is to allow the users to transmit asynchronously (even though they may have a synchronized clock!) in the sense that they transmit as soon as they have a message. The encoded message is spread over a time interval comparable to D to ensure reliable communication.

Chapter 3 shows that random coding can achieve the asynchronous capacity for this channel model. Each user would have to transmit a preamble to identify himself as the current user of the channel. For the converse, we show that reliable communication is impossible above certain asynchronous sum-capacity, which will be defined in chapter 3.

1.4 The issue of feedback

The multiple access channel with feedback is modeled in figure 1.4.1. For the classical one way memoryless channel, it is well known that feedback cannot improve channel capacity. For multiple accessing, it has been shown that feedback improves the capacity region of the two-user adder channel [1], the additive Gaussian noise channel [1], as well as increases the throughput of the tree algorithm beyond the e^{-1} capacity for the collision channel without feedback. It is important to understand how feedback affects the operation of the multiple access channel, and why the capacity region of the multiple access channel may be enlarged by feedback.

There are three plausible reasons for the enlargement of the capacity region in the presence of feedback. First, the use of feedback may improve synchronization among the transmitters so that signaling interference among the users can be minimized. Second, feedback may enable us to achieve a probability distributions for the channel output that is not achievable by independent probability distributions for the inputs of the users. Third, feedback may provide information about the set of simultaneous users, thus helping to achieve better scheduling during retransmission.

One example that illustrates the first two reasons is the two-user noiseless adder channel of Wolf, with its capacity region shown in figure 1.4.2. With full cooperation between the

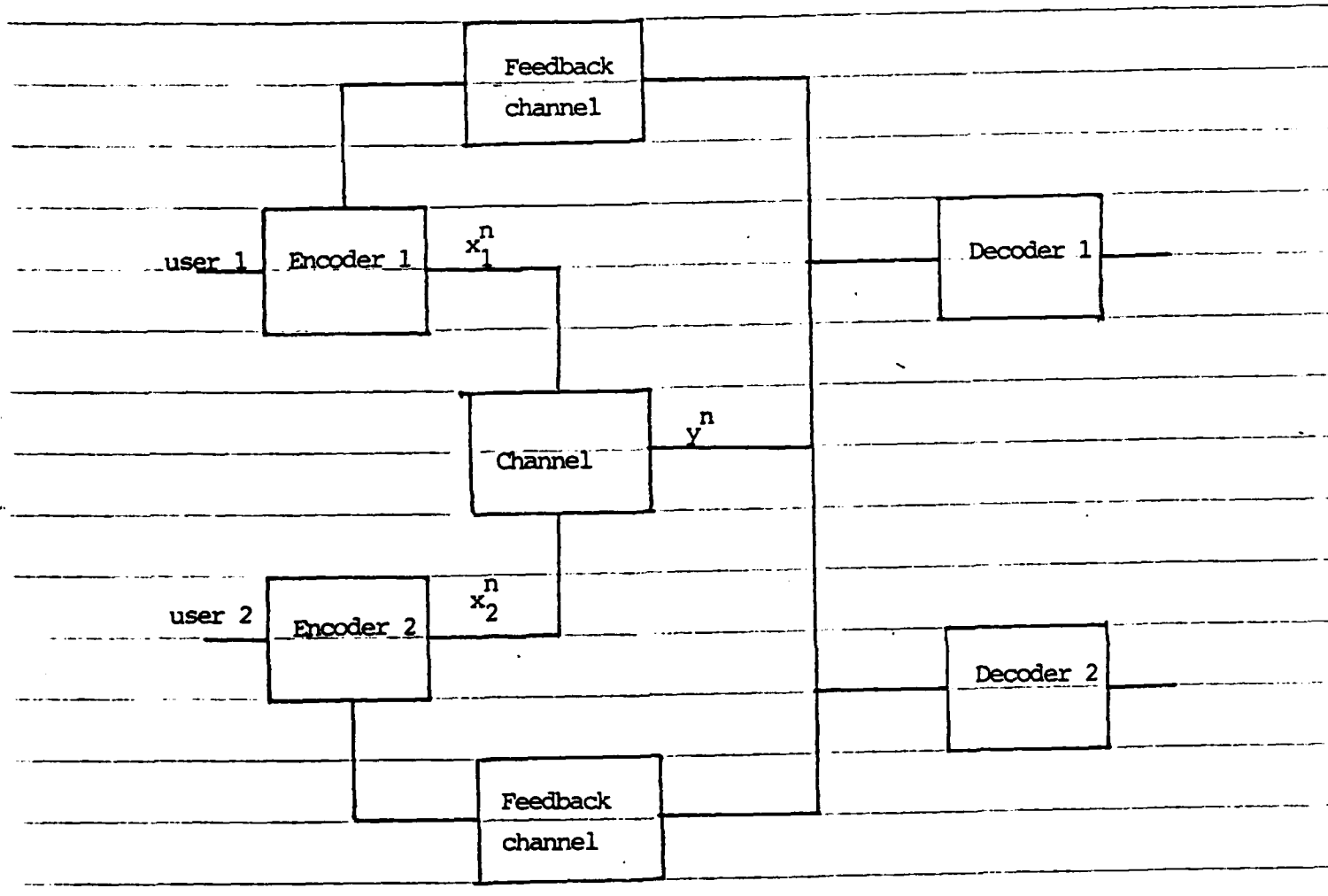


Figure 1.4.1 The multiple access channel with feedback

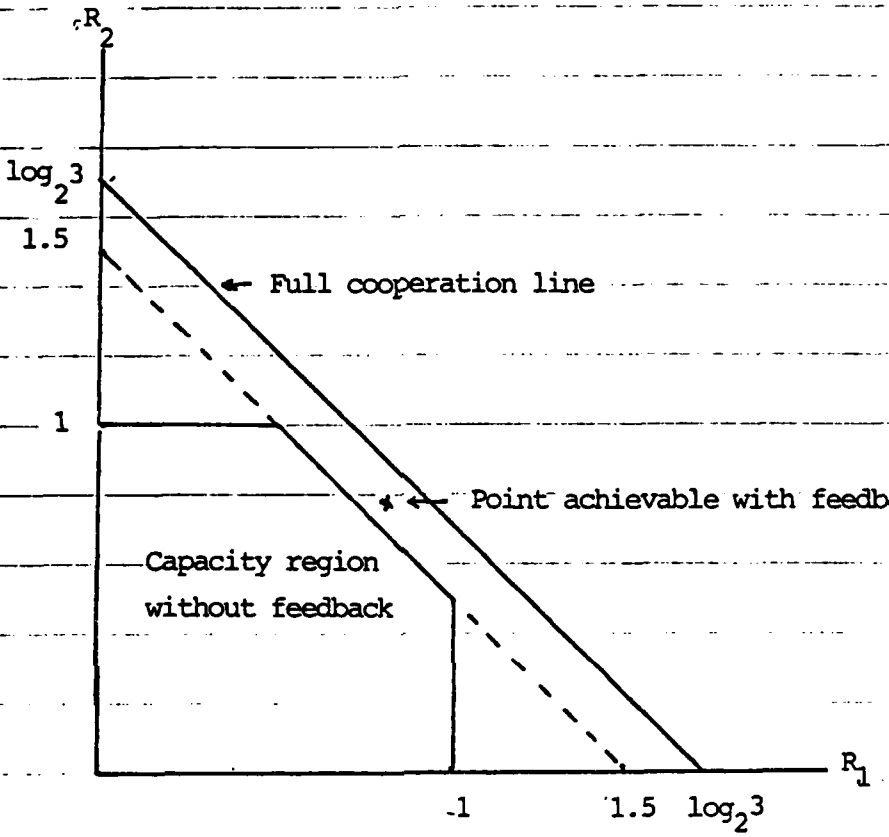


Figure 1.4.2 The capacity region of the two-user adder channel

two users, the capacity of $\log_2 3$ bits per symbol can be achieved by using the letters in the three level output alphabet $\{-2,0,2\}$ equiprobably. In the absence of cooperation and feedback, an equiprobable output alphabet cannot be achieved. The best throughput is achieved by having each user transmitting independently, with equal probability distribution for the input alphabet $\{-1,1\}$. The resulting output distribution is then $P(y=-2) = P(y=2) = 1/4$ and $P(y=0) = 1/2$. Consequently, the entropy of Y is 1.5 bits per symbol. The channel throughput without cooperation, which equals the entropy of Y in this case, is less than the throughput of $\log_2 3$ bits for the case with full cooperation. Consider now the feedback of y to the two transmitters. By observing y, the first transmitter can derive the code symbol of the second transmitter simply by subtracting from y the symbol the first transmitter sent. Similarly, the second transmitter can derive the code symbol sent by the first transmitter. Therefore, locations where there are ambiguities (to the receivers) about the symbols sent by the transmitters (that is, where the channel output is 0) may be resolved cooperatively by retransmission. Hence the transmitters can cooperate fully in resolving these ambiguities during retransmission, thus achieving a higher throughput for the channel. The two transmitters must be synchronized during retransmission to achieve full cooperation. The observation of the channel output may help to synchronize the users.

If feedback does not help to synchronize the retransmission (or it is ignored for the purpose of

synchronization), then the capacity region cannot be enlarged by feedback. For the Aloha scheme, retransmissions are not synchronized among the users, which explains why its capacity is the same as that of the collision channel without feedback.

Neither the tree algorithm nor the Aloha scheme explicitly require a synchronized clock for each user. For the Aloha scheme, feedback can be asynchronous in the sense that feedback about occurrences of collisions can have randomly large delay without affecting the throughput. The tree algorithm, in contrast, requires synchronized feedback, which helps to achieve a certain degree of synchronism for the users once a conflict arises. Therefore, the capacity of the tree algorithm is larger than that of the Aloha scheme.

Feedback can also provide information about the set of simultaneous users. For the Aloha scheme, a collision informs the transmitter that there is at least one other message transmitted simultaneously. Other transmitters need not be provided this information. For the tree algorithm, feedback distinguishes the cases of none, one or more than one message in a slot. This information is broadcast to every transmitter. Such information about the size of the set of simultaneous users improves channel capacity.

Besides these three reasons why feedback may improve throughput, feedback very often reduces the complexity of encoding for reliable communication. The Aloha scheme and the tree algorithm do not require error correction coding for their

operation, whereas error correction coding is indispensable for the collision channel without feedback. The doubly exponential error exponent of the multiple access OR channel with feedback has been established by Györki and Kerekes [8].

It is hard to obtain closed form characterization of channel capacity for the multiple access channel with feedback. Therefore, the issue of feedback will not be given further attention in this thesis.

1.5 The issue of codebook knowledge

If the codebooks of each transmitter are known at the receivers, each receiver may determine jointly the most likely code sequences sent by the transmitters. However, this joint decoding effort is often computationally infeasible. Consequently, the signals of the other users are often treated as interfering noise which are not estimated for noise reduction. The code division multiple access scheme for the spread spectrum channel is one example in which joint decoding is not performed. Incomplete codebook knowledge also arises from two other situations. First, some of the codebooks of the other users may not be known to a receiver for security reasons. Second, jamming may occur in the system. The capacity of these channels with incomplete knowledge of the codebooks is of great practical interests. The characterization of the capacity region for such channels is the subject of chapter 4.

As an illustration, consider a multiple access channel with two users U_X and U_Z. Each codeword x^n and z^n used by U_X and U_Z respectively are required to satisfy the following constraints

$$c_X(x^n) = \sum_{k=1}^{n} c_X(x_k^n) \leq nC_X$$

$$c_Z(z^n) = \sum_{k=1}^{n} c_Z(z_k^n) \leq nC_Z$$

in which the subscript k denotes the k-th symbol of a sequence; and c_X , c_Z are cost functions on the alphabets X and Z respectively. Suppose the receiver of U_X only knows that the

codewords of U_Z satisfies the second constraint. The question then is how the receiver of U_X should perform decoding.

More specifically, we assume that the decoder of U_X uses some metric $a_{xy} \geq 0$ for each $x \in X$ and $y \in Y$ (Y is the channel output alphabet). The decoding rule is to choose the message with a codeword that maximizes the sum of the metric between each symbol in the codeword and the corresponding channel output symbol. The choice of the metric a_{xy} depends on the decoder's knowledge (which can be incorrect) about the channel or the other users of the channel. Thus by imposing a specific structure for the decoder, we try a novel approach to the problem of robust decoding. This approach generates three results in chapter 4.

The first result gives an achievable rate for a single user U_X communicating over a stationary and memoryless channel with channel transition probabilities p(y/x). The decoder uses a metric $\{a_{xy}\}$ which may be different from that used for maximum likelihood decoding, namely $a_{xy} = \ln p(y/x)$. The result states that reliable communication is achievable for rates less than

$$C = \max_{P(X)} I'(X;Y)$$

in which

$$I'(X;Y) = H(X) + H(Y) - H'(XY)$$

$$H'(XY) = \max_{f_{xy}} H(\{f_{xy}\}) = \max_{f_{xy}} \sum_{xy} -f_{xy} \ln f_{xy}$$

subject to

$$f_{xy} \geq 0 \quad \text{for all } x, y$$

$$\sum_{x,y} a_{xy} f_{xy} \geq \sum_{x,y} a_{xy} p_{xy};\ p_{xy} \triangleq P(X=x, Y=y) = P(X=x).P(Y=y/X=x)$$

$$\sum_{y} f_{xy} = \sum_{y} p_{xy} \quad \text{for all } x$$

$$\sum_{x} f_{xy} = \sum_{x} p_{xy} \quad \text{for all } y$$

The function $I'(X;Y)$ has the pleasing property of

$$0 \leq I'(X;Y) \leq I(X;Y)$$

Furthermore $I'(X;Y)$ is convex $\cup$ in $P(Y/X)$. We conjecture that C is also an upper bound for the achievable rate, thus making C the capacity of the channel with given decoder structure.

The second result involves the two-user channel stated at the beginning of this section, with U_Z trying to jam U_X. Let $\{a_{xy}\}$ be the metric used by the decoder of U_X. Define the capacity C for U_X (which has a given decoder structure) as the maximum rate that U_X may transmit reliably for all codewords of U_Z satisfying the constraint

$$c_Z(z^n) \leq nC_Z$$

The result states that

$$C = \min_{P(Z)\in\mathcal{Z}} \max_{P(X)\in\mathcal{X}} I(X;Y) \qquad \text{.....*}$$

in which $\mathcal{Z}$ is the set of probability distributions on the alphabet Z satisfying

$$\sum_{z \in Z} P(z)\, c_Z(z) \leq C_Z$$

and $\mathcal{X}$ is the set of probability distributions on the alphabet X satisfying

$$\sum_{x \in X} P(x)\, c_X(x) \leq C_X$$

This capacity is achieved when a_{xy} is chosen to be ln P(xy) with

$$P(xy) = \sum_z P(xyz) = \sum_z P^*(x)\, P^*(z)\, P(y/xz)$$

in which $P^*(X)$ is the probability distribution that achieves the maximum in *, and $P^*(Z)$ is the probability distribution that achieves the minimum in *.

The third result involves a system with M asynchonous users. There is an (M+1)-th source, which tries to jam the M users. We assume that receiver m is only interested in the message of user m. The receiver m knows completely the codebooks of those users in the index set $I_m \subseteq \{1,2,\ldots,M\}$. The receiver m performs joint decoding for the set I_m, while assuming a metric which takes into account the effect of the channel and the users in $\bar{I}_m = \{1,2,\ldots,M\} - I_m$. A set of constraints

$$c_i(x_i^n) = \sum_{k=1}^{n} c\,(x_{i,k}^n) \leq nC_i$$

for all $1 \leq i \leq M+1$ are placed on the codewords x_i^n of the M users and the jammer. The result states that the capacity region $\mathcal{R}$ is given by

$$\mathcal{R} = \bigcap_{P(X_{M+1}) \in \mathcal{X}_{M+1}} \text{convex hull} \bigcup_{P(X_i) \in \mathcal{X}_i, 1 \leq i \leq M} \tilde{\mathcal{R}}$$

in which $\mathcal{X}_i$ is the set of random variables X_i satisfying

$$\sum_{x_i \in X_i} P(x_i)\, c_i(x_i) \leq C_i$$

for $1 \leq i \leq M+1$ and $(R_1, \ldots, R_M) \in \tilde{\mathcal{R}}$ iff

$$\sum_{i \in \Omega_m} R_i < I(\{X_i\}_{i \in \Omega_m}; Y / \{X_i\}_{i \in \bar{\Omega}_m})$$

for all $1 \leq m \leq M+1$ and for all

$$\Omega_m \subseteq I_m \quad \text{and} \quad m \in \Omega_m$$

$$\bar{\Omega}_m = I_m - \Omega_m$$

Thus for the two-user, no jammer asynchronous multiple access channel without joint decoding ($I_1=\{1\}$, $I_2=\{2\}$), the capacity region $\mathcal{R}$ is given by

$$\mathcal{R} = \bigcup_{P_1(X_1), P_2(X_2)} \tilde{\mathcal{R}}$$

in which $(R_1, R_2) \in \tilde{\mathcal{R}}$ if

$$R_1 < I(X_1; Y)$$

$$R_2 < I(X_2; Y)$$

A specific $\tilde{\mathcal{R}}$ is shown in figure 4.5.1, which is compared with an $\tilde{\mathcal{R}}$ for the case with joint decoding.

1.6 Specific channels and coding schemes

The general theories stated in the previous sections will be applied to three channels of practical interest, namely the OR channel, the spread spectrum channel and the collision channel. The number of users is large, with only a small portion active at a time. Due to the difficulties of scheduling a large number of users as well as maintaining synchronization, the users will transmit asynchronously. Joint decoding is considered for the OR channel. For the collision channel, there is hardly any distinction between the cases with and without joint decoding. The asynchronous capacity region for M active users will be derived. The point of most interest in the capacity region is where the rates of the users are equal and maximized (that is, the farthest point on the main diagonal that is inside the capacity region). The capacity of these multiple access channels, defined as the sum of the rates at the point of interest, is derived for large M.

Specific code division multiple access schemes are proposed for these channels. Such schemes have three aspects. The first aspect involves the type of modulation used (PPM or on-off keying for the OR channel; packeting, superpacketing and interleaving for the collision channel; and the use of PN sequence for the spread spectrum channel). The second aspect involves the type of code used (block, convolutional, rate, constraint length or block size). The third aspect involves the type of decoding algorithm used (sequential or Viterbi).

All three aspects are important for determining three performance measures, namely, error probability, decoding complexity and maximum achievable throughput, which can be different from the capacity of the multiple access channel. These three performance measures are given extensive treatment in chapters 5, 6 and 7. An outline of the general approach will be given here.

In general, we prefer the use of convolutional codes to block codes for their superior error performance and relative ease of decoding. The code rate used is typically low (1/2, 1/3 or even smaller) because interference from other users can be quite severe. Severe interference also necessitates the use of long constraint length codes to improve error performance.

An upper bound on throughput, which is of more interest than capacity when sequential decoders are used, is the cutoff rate R_o, which is the maximum rate of encoding to guarantee a bounded average computation per information bit for sequential decoding. The cutoff rate is computed for the three channels considered, with the interference of the other users treated as memoryless noise. It is noteworthy that cutoff rate, unlike capacity, depends on the modulation used. (In fact, PPM has a higher cutoff rate than on-off keying for the OR channel.) The cutoff rate is used for two purposes. First, the bit error rate P_b for the random code ensemble can be conveniently upper bounded as a function of the constraint length and the rate of the convolutional code, as well as the cutoff rate. We expect the

bit error prababilities for specific codes to agree closely with this random coding bound since for long constraint length codes, we have no better way to select good codes than to pick them at random. Since the constraint length is long, we expect low probability of error due to path merging [10] in the trellis. The second and more important use of the cutoff rate is for upper bounding the decoding complexity $\bar{C}_j$ defined as the average computation per node for sequential decoding over the random code ensemble. The bound shows that average computation per node for the OR channel or the collision channel is small (less than 10) for throughput close to the cutoff rate.

The obvious question then is the maximum sum of the throughputs for all users that can be reliably achieved by convolutional codes with sequential decoding. Let T be the total throughputs of a specific scheme, with each user employing a rate r convolutional code. Given the information rate (T/M) each user wants to convey and the method of encoding the information, the level of interference contributed by the user to the channel is determined. Hence the cutoff rate for each user is a function $R_0(T,r)$ of T and r. In order that reliable communication may be achieved with sequential decoding, the rate r must be less than the cutoff rate, hence $r < R_0(T,r)$. For given r, we solve for the largest T which satisfies the above inequality. We may subsequently maximize T over convenient values of r. The result is the maximum sum of the throughputs for all users.

Expressions in the form

$$P_b = P_b(R_o, k, r) = P_b(T, k, r)$$
$$\bar{C}_j = \bar{C}_j(R_o, r) = \bar{C}_j(T, r)$$

are also obtained for the three channels. The values of P_b and $\bar{C}_j$ versus T, for different values of k and r are plotted in the thesis. The probability of buffer overflow for sequential decoding is also examined for the OR channel and the collision channel.

Several specific issues concerning these channels arise in chapters 5,6 and 7, and a few highlights are listed here before we conclude this discussion.

1. For the spread spectrum channel, using a short PN sequence and a low rate convolutional encoder achieves a higher throughput (.721 for capacity and .361 for cutoff rate) than using a long PN sequence and a high rate convolutional encoder.

2. For the OR channel, PPM achieves a higher total cutoff rate (.69) than on-off keying (.59). We recommend using a single bit input, single PPM symbol output (that is, rate = 1 bit/1 PPM symbol) convolutional code. A forward search decoding algorithm described in chapter 5 requires very low complexity even at a high total throughput of .5. Nonbinary trellises should not be used because they have lower throughput.

3. For the noiseless collision channel, we recommend using a rate 1/3 convolutional encoder and a forward search decoding algorithm. The maximum sum of throughput is .295. The decoding complexity is low even at a high throughput of .28.

4. Consider the collision channel corrupted by Gaussian noise, in addition to erasure 'noise' due to collisions. Chapter 7 shows that substantial power saving is gained, at hardly any extra cost, by the redundant coding that is originally intended to correct erasures due to packet collisions. A signal to noise ratio of 4 or 5 dB would be sufficient, which compares favourably with the typical 10.5 dB required for uncoded antipodal signaling for a bit error rate of 10^{-6} .

Chapter 2 The Asynchronous Multiple Access Channel

The main purpose of this chapter is to provide an understanding of the effect of asynchronism on multiple accessing. We shall assume that the codebooks for the users are known to every receiver. Section 2.1 gives a precise definition of asynchronism for the M-user multiple access channel and characterizes the asynchronous capacity region. Sections 2.2 and 2.3 prove the direct part and the converse of the coding theorem respectively for the two-user case. The proof of the theorem for the M-user case is a straight forward extension of the result for the two-user case. The subject of incomplete codebook knowledge is studied in chapter 4. The characterization and proof for the capacity region of the asynchronous M-user multiple access channel with incomplete codebook knowledge are conceptually straight-forward once the essence of the proofs in chapter 2 and 4 is well understood.

2.1 Modeling and characterization of the capacity region

Let there be M sources sending independent messages. The source i, $1 \leq i \leq M$ uses a codebook C_i containing 2^{nR_i} equiprobably used codewords each representing the message $j_i \in J=\{1,2,\ldots,2^{nR_i}\}$. Each codeword $x_i(j_i)=(x_{ik}(j_i))$, $1 \leq k \leq n$, where $x_{ik}(j_i) \in X$, the signaling alphabet. Let C be the ordered M-tuple (C_i), $1 \leq i \leq M$.

We shall look at how asynchronism arises in communication systems. Each transmitter has a clock that runs accurately. However, the initialization may be different for each clock. Suppose a standard clock exists. Let Δ_i be the amount of time the clock i is running ahead of the standard clock ($\Delta_i < 0$ for a lag). We may have some a priori knowledge about Δ_i, such as upper and lower bounds or more specifically, a joint probability density $P(\Delta_1, \Delta_2, \ldots, \Delta_M)$ which we shall assume known henceforth. We assume that the Δ_i's are mutually independent, thus $P(\Delta_1, \Delta_2, \ldots, \Delta_M) = P(\Delta_1).P(\Delta_2)\ldots P(\Delta_M)$. Since the channel is assumed to be symbol synchronous, time may be considered as discrete and hence the Δ_i's may be assumed to be integers.

Communication systems often have a fuzzy agreement of time. While the transmitters may not be synchronized up to microseconds, they should have an idea of the hour or the day. Hence $P(\Delta_i)$ is usually nonzero only over a finite interval. Let

$$\Delta_{max} = \max_{i,j} \max \{ |\Delta_i - \Delta_j| : P(\Delta_i), P(\Delta_j) > 0 \}$$

Consider the use of round-robin TDMA for the collision channel with guard time Δ_{max} between consecutive transmission intervals of two different users. If the transmission interval of a user is much larger than Δ_{max}, the amount of overhead due to guard time is negligible. Therefore full channel capacity may be achieved. The penalty, however, is a large delay and increased buffer requirement. Thus, the encoder may be constrained to have a

complexity (including buffering) that falls far short of that required to achieve full channel capacity using round-robin TDMA. Since this encoder with "limited complexity" operates over a long time interval, the encoder would have to be used "periodically". We would like to make the terms in quotes precise through the following formulation. The complexity of the code is the length of the codewords, namely n. The codebooks are used periodically. Each source sends the codewords $x_i(j_i^t)$ sequentially, where t is an integer and j_i^t is the t-th message sent by the user i (figure 2.1.1). The codeword $x_i(j_i^t)$ starts at time 0 registered by the clock at the transmitter i. Time for each transmitter is divided into periods of length n. The random variable D_i is the delay between the start of periods for the standard clock and the transmitter i (figure 2.1.1). Thus $0 \leq D_i \leq n-1$. Consequently, $P(D_i)$ is formed by aliasing $P(\Delta_i)$, or

$$P(D_i = d_i) = \sum_{t=-\infty}^{\infty} P(\Delta_i = nt + d_i)$$

The degree of asynchronism is totally specified by the knowledge of n and $P(D_i)$, $1 \leq i \leq M$. We say that a system is perfectly synchronous if $P(D_i=0)=1$ for all $1 \leq i \leq M$. A system is uniformly asynchronous if $P(D_i)$ is uniform for $0 \leq D_i \leq n-1$. A system is very asynchronous if it is uniformly asynchronous for large n. In essence, a very asynchronous system has the clock at each transmitter randomly initialized over a long period of time.

However, imposing an upper bound on the complexity measure n of the encoder may cause some problems when we apply classical

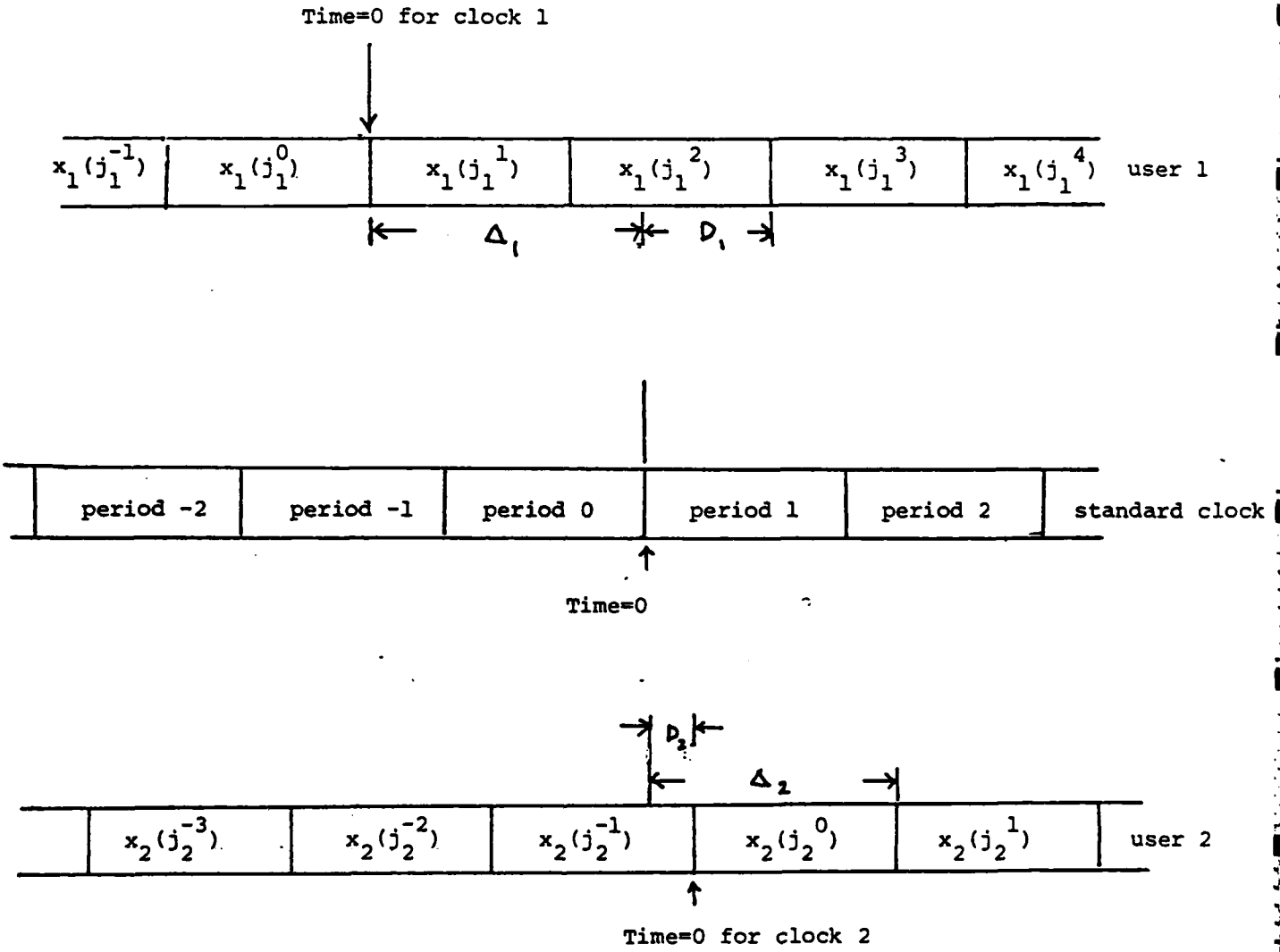


Figure 2.1.1 Code asynchronism

information theory. In proving the direct part of coding theorems, the error probability approaches zero as the complexity measure n increases without bound. Thus, a finite encoder complexity may give a nonzero error probability. However, we shall assume that the upper bound n is sufficiently large so that the error probability would be satisfactorily small. Yet n is not large enough for using round-robin TDMA that achieves full channel capacity. For systems that are very asynchronous, the upper bound n can be very large, in which case asymptotical zero error probability may be approached.

The channel is assumed to be memoryless with $P(y/x_1 x_2 \ldots x_M)$ known for all $x_i \in X$, $1 \leq i \leq M$ and $y \in Y$. We shall assume that the receiver ℓ is only interested in the messages sent by the transmitter ℓ. The decoder observes the channel output sequence and concludes that the sent messages are $\hat{j}_\ell^\tau$.

One performance measure of a coding system is the probability of bit error for each receiver, which is defined as

$$\langle P_{e,\ell}^{t} \rangle = 1/nR_\ell \sum_{k=1}^{nR_\ell} P(\hat{J}_{\ell k}^{t} \neq J_{\ell k}^{t})$$

in which $\hat{J}_{\ell k}^{t}$ and $J_{\ell k}^{t}$ are respectively the k-th bit of the binary representation of $\hat{J}_\ell^\tau$ and J_ℓ^t. The capacity region $\mathcal{R}$ is defined as the set of all M-tuples $(R_1, R_2, \ldots, R_M)$ such that there exist sequences of codebooks for each source which make $\langle P_{e,\ell}^{\tau} \rangle$ go to zero for all ℓ and t as n increases. Define the region $\bar{\mathcal{R}}$ by

$(R_1,\ldots,R_M) \in \bar{\mathcal{R}}$ iff for some

$$P(x_1,x_2,\ldots,x_M,y) = P_1(x_1).P_2(x_2)\ldots P_M(x_M).P(y/x_1,\ldots x_M),$$

$$\sum_{i\in\Omega_\ell} R_i < I(\{X_i\}_{i\in\Omega_\ell};\ Y/\ \{X_i\}_{i\in\bar{\Omega}_\ell}),\ 1\leq\ell\leq M$$

for all $\Omega_\ell \subseteq \{1,2,\ldots,M\}$ and $\ell \in \Omega_\ell$

and $\bar{\Omega}_\ell = \{1,2,\ldots,M\} - \Omega_\ell$

The major result of this chapter is that for the very asynchronous channel, $\mathcal{R} = \bar{\mathcal{R}}$. For the two-user case, $\bar{\mathcal{R}}$ is given by

$(R_1,R_2) \in \bar{\mathcal{R}}$ iff for some

$$P(x_1,x_2,y) = P_1(x_1).P_2(x_2).P(y/x_1x_2)$$

$$R_1 < I(X_1;\ Y/\ X_2)$$

$$R_2 < I(X_2;\ Y/\ X_1)$$

$$R_1 + R_2 < I(X_1X_2;\ Y)$$

In contrast to the capacity region for the synchronous multiple access channel, the above characterization does not contain a convex hull operation. In essence, the time sharing argument for forming the convex hull of the capacity region is not permitted since the two users are asynchronous. In a paper studying the asynchronous multiple access channel, Cover et. al. [7] argue that the capacity region is not affected by asynchronism. However, the asynchronism they considered has a code complexity n which is much larger than the maximum delay. In contrast, the

very asynchronous channel we are studying involves codes with complexity much smaller the the maximum delay.

Section 2.2 proves the direct part ($\bar{\mathcal{R}} \subset \mathcal{R}$) for the two-user case by achieving reliable communication for all points in $\bar{\mathcal{R}}$ through random coding. Section 2.3 proves the converse ($\mathcal{R} \subset \bar{\mathcal{R}}$) by showing that all codes with rate outside of $\bar{\mathcal{R}}$ has a bit error probability lower bounded above zero.

2.2 Direct part for the two-user asynchronous channel

The proof of the direct part employs random codes to achieve all points of $\bar{\mathcal{R}}$. The set of codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$ for the two users are generated randomly in the following manner. Each codebook $C_i \in \mathcal{C}_i$, $i = 1,2$ contains the codewords $x_i(j_i)$, each chosen independently for the message $1 \leq j_i \leq 2^{nR_i}$. Each letter $x_{ik}(j_i)$ in the codeword is chosen independently for $1 \leq k \leq n$, according to the probability distribution $P_i(X)$. Therefore

$$P((X_i(j_i) = x_i) = \prod_{k=1}^{n} P(X_{ik}(j_i) = x_{ik})$$

For the sake of simplicity for the proof, the encoder has two special features. First, the encoder inserts a preamble once every m messages for the purpose of code synchronization. The preambles used in the proof are the codewords $x_i(J_i=1)$ or $x_i(J_i=2)$, which are set aside for the purpose of code synchronization. The preambles are recognized by special devices at the decoder, and consequently give the values of the delays d_1 and d_2. A description of this code synchronization mechanism, a justification of the fact that code synchronization can be achieved and a quantitative estimate of the amount of preamble overhead required to achieve synchronization are given in Appendix 2.1.

The second special feature is that the encoders put out codewords in such a way that in any sequence of m+1 consecutive messages (including the preambles), none of the messages is repeated. In other words, $J_i^t \neq J_i^{t'}$ for all $t \neq t'$ and $0 \leq t, t' \leq m+1$.

The encoder achieves this by keeping a record of the past m messages. If the next message has occured in the past m messages, it is put aside for a while and the next message is considered. Provided $m << 2^{nR_i}$, the backlog of messages should be sufficiently small compared with m. Since there is no repetition in m+1 consecutive messages, any two disjoint sets of code symbols in any m+1 consecutive codewords for both users would be statistically independent over the random code. This independence property will be used in proving the direct part of the coding theorem. Other than for the purpose of proving the direct part, the insurance that the messages are not repeated seems unnecessary for reliable communication.

The decoding for the first receiver is performed as shown in figure 2.1.1. The first receiver detects the preambles $x_1(J_1^{0}=1)$ and $x_1(J_1^{m+1}=2)$ for the first user, and $x_2(J_2^{t}=1)$ for some $0 \leq t \leq m$ for the second user. (In figure 2.2.1, t = 2.) For the sake of independence, we have used two different preamble sequences alternately. It is assumed that the preamble sequences $x_1(J_1^{0}=\cdot)$, $x_1(J_1^{m+1}=\cdot)$, $x_2(J_2^{t}=\cdot)$ as well as the values D_1 and D_2 are detected correctly. This assumption is justified in Appendix 2.1. Without loss of generality, D_1 is assumed to be larger than D_2. The decoder for the first user decodes by observing Y^{r}, which starts as $x_2(J_2^{0})$ begins transmission and ends as $x_2(J_2^{m})$ terminates transmission, as shown in the figure. Hence the length of Y^{r} is $r=(m+1)n$. The corresponding code symbols for user 1 and user 2 are denoted respectively as $x_1^{+}(J_1^{(m)})$ and $x_2(J_2^{(m)+})$, where

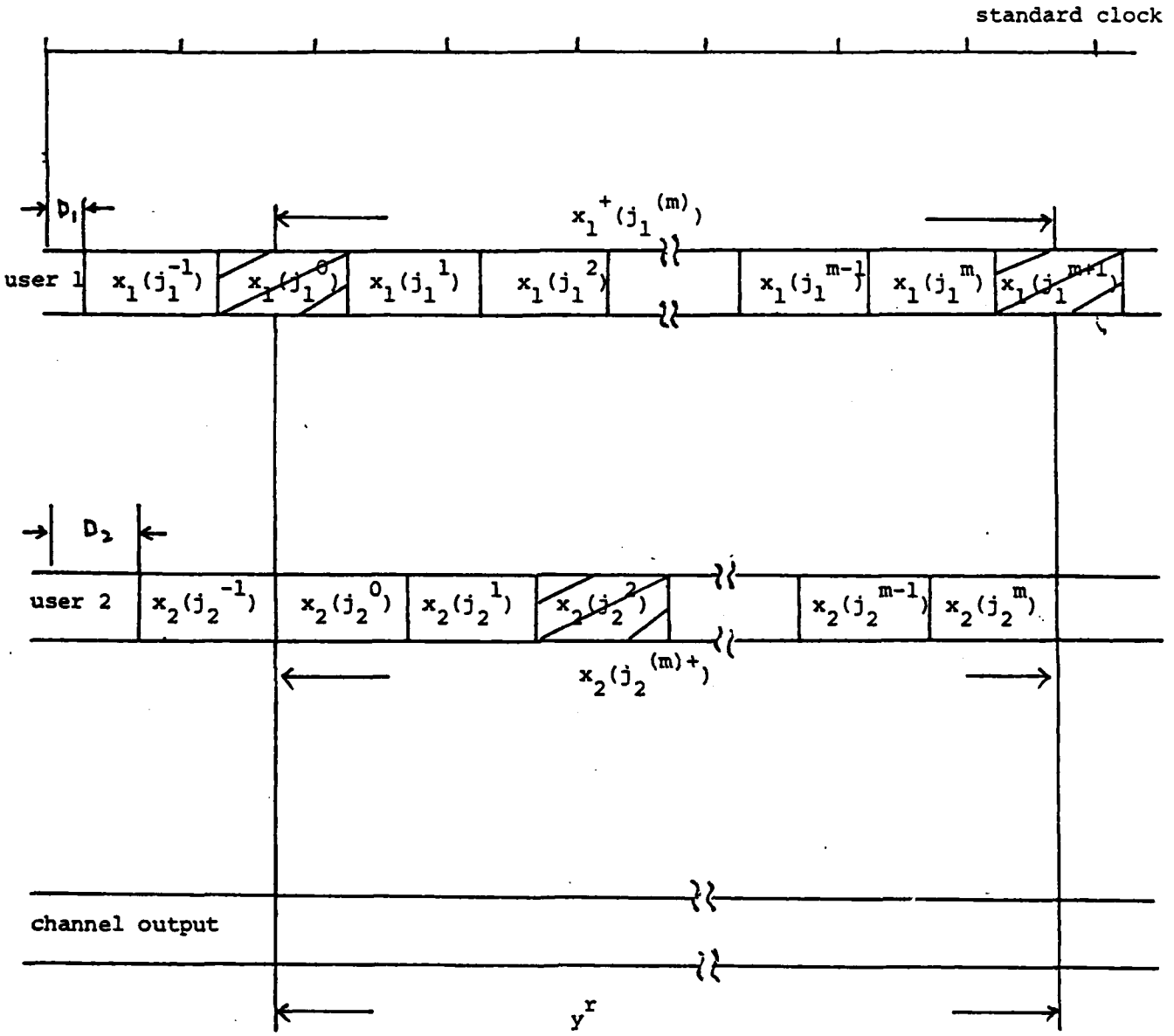


Figure 2.2.1 Decoding scheme

$$J_1^{(m)} = (J_1^1, J_1^2, \ldots, J_1^m)$$

$$J_2^{(m)+} = (J_2^0, J_2^1, \ldots, J_2^m)$$

There is a special reason why the channel sequence Y^τ is blocked in such a manner for decoding the message of the first user. The estimation of j_1^t depends on the estimation of j_2^τ as well as $j_2^{\tau-1}$ since $x_1(j_1^t)$ overlaps with $x_2(j_2^\tau)$ and $x_2(j_2^{t-1})$. The estimation of j_2^{t-1} in turn depends on the estimation of $j_2^{t-2}, j_2^{\tau-1}$ and j_2^τ; etc. Thus, the estimation of the messages is chained. This chain is terminated when the estimation of j_2^0 (or at the other end j_2^m) depends on the estimation of j_1^1 only (or j_1^m only) since the preamble j_1^0 (or j_1^{m+1}) is known. Thus the preambles, besides synchronization, also eliminate a "boundary effect" in the process of estimation.

The decoding function for the first user is defined by

$$g_1(y^\tau) = \hat{j}_1^{(m)} = (\hat{j}_1^1, \hat{j}_1^2, \ldots, \hat{j}_1^m) = \text{any element of } G_1(y^\tau)$$

in which

$$G_1(y^\tau) = \{(\tilde{j}_1^1, \tilde{j}_1^2, \ldots, \tilde{j}_1^m): (x_1^\tau(\tilde{j}_1^{(m)}), x_2(\tilde{j}_2^{(m)+}), y^\tau) \in T_\epsilon^\tau\}$$

We shall give a definition of T_ϵ^τ later on. If $G_1(y^\tau)$ is empty, then the decoder signals a decoding failure. Before giving a definition of T_ϵ^τ, we would like to examine $P(x_1^\tau(\tilde{j}_1^{(m)}), x_2(\tilde{j}_2^{(m)+}), y^\tau / j_1^{(m)}, j_2^{(m)+})$. Since the codewords are generated independently symbol by symbol and used nonrepeatedly, and the channel is memoryless, we have

$$P(x_1^+(\tilde{j}_1^{(m)}), x_2(\tilde{j}_2^{(m)+}), y^r / j_1^{(m)}, j_2^{(m)+})$$

$$= \prod_{k=1}^{(m+1)n} P(x_{1k}^+(\tilde{j}_1^{(m)}), x_{2k}(\tilde{j}_2^{(m)+}), y_k^r / j_1^{(m)}, j_2^{(m)+})$$

in which the subscript k denotes the k-th symbol of a sequence. Four cases (figure 2.2.2) can arise at the k-th location (which is assumed to be embedded in the t-th codeword) of the three sequences $x_1^+(j_1^{(m)}..)$, $x_2(j_2^{(m)+})$ and y^r,

1. $\tilde{j}_1^t = j_1^t$
   $\tilde{j}_2^t = j_2^t$
   For this case, we would leave
   $P(x_{1k}^+(\tilde{j}_1^{(m)}), x_{2k}(\tilde{j}_2^{(m)+}), y_k^r / j_1^{(m)}\ j_2^{(m)+})$
   as it is in the product.

2. $\tilde{j}_1^t \neq j_1^t$
   $\tilde{j}_2^t = j_2^t$
   since $\tilde{j}_1^t \neq j_1^t$ and different codewords are generated independently,
   $$P(x_{1k}^+(\tilde{j}_1^{(m)}), x_{2k}(\tilde{j}_2^{(m)+}), y_k^r / j_1^{(m)}\ j_2^{(m)+})$$
   $$= P(x_{1k}^+(\tilde{j}_1^{(m)})) \cdot P(x_{2k}(\tilde{j}_2^{(m)+}), y_k^r)$$

3. $\tilde{j}_1^t = j_1^t$
   $\tilde{j}_2^t \neq j_2^t$
   Similar to 2
   $$P(x_{1k}^+(\tilde{j}_1^{(m)}), x_{2k}(\tilde{j}_2^{(m)+}), y_k^r / j_1^{(m)}, j_2^{(m)+})$$
   $$= P(x_{1k}^+(\tilde{j}_1^{(m)}), y_k^r) \cdot P(x_{2k}(j_2^{(m)+}))$$

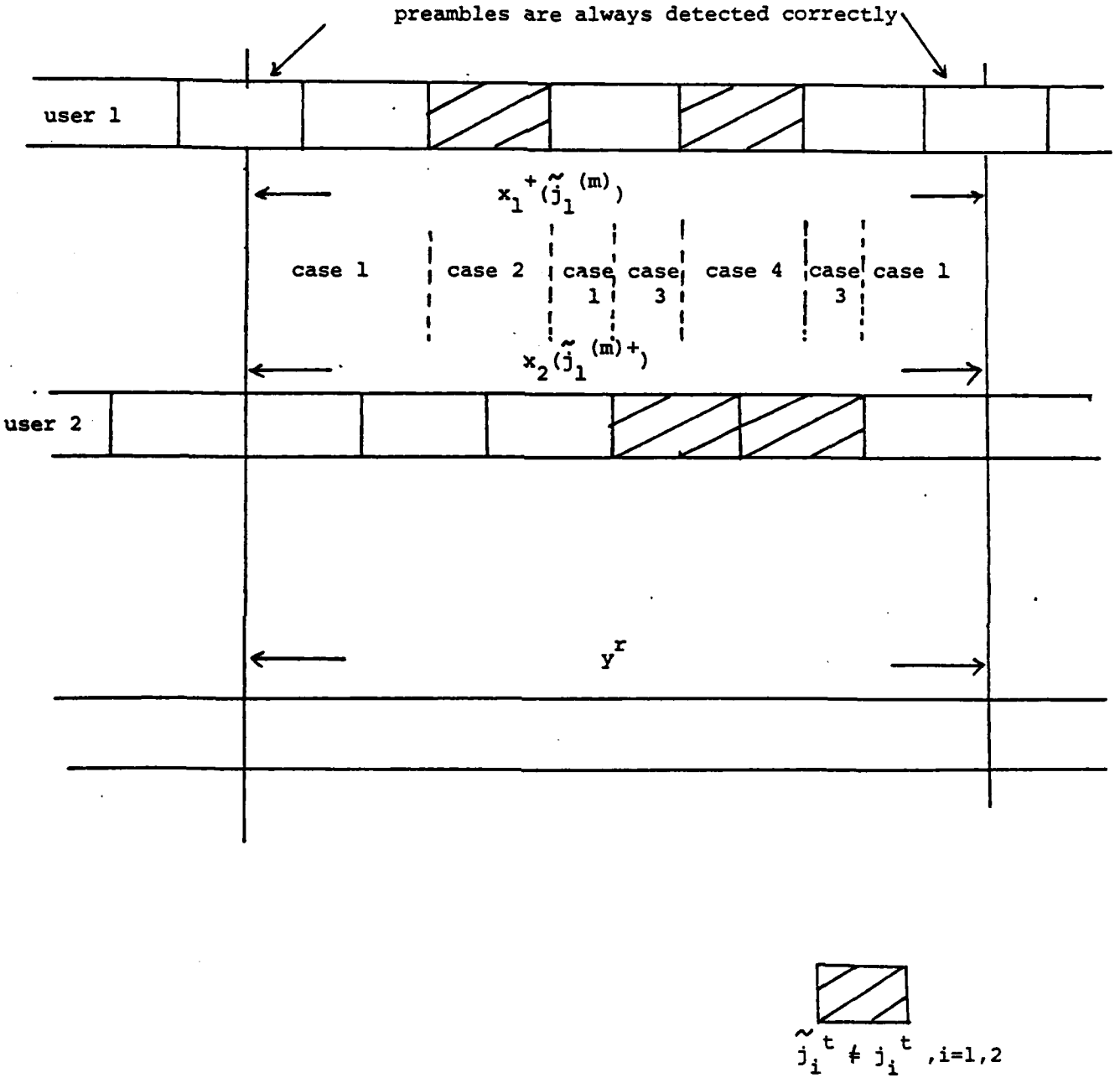


Figure 2.2.2 Four types of errors

4. $\tilde{j}_1^t \neq j_1^t$

$\tilde{j}_2^t \neq j_2^t$

in which case

$$P(x_{1k}^{+}(\tilde{j}_1^{(m)}), x_{2k}(\tilde{j}_2^{(m)+}), y_k^r / j_1^{(m)}, j_2^{(m)+}) = P(x_{1k}^{+}(\tilde{j}_1^{(m)})) \cdot P(x_{2k}(\tilde{j}_2^{(m)+})) \cdot P(y_k^r)$$

For the four cases mentioned above, the expected value of $-\log P(x_{1k}^{+}(\tilde{j}_1^{(m)}), x_{2k}(\tilde{j}_2^{(m)+}), y_k^r / j_1^{(m)}, j_2^{(m)+})$ are respectively $H(X_1X_2Y)$, $H(X_1)+H(X_2Y)$, $H(X_1Y)+H(X_2)$, $H(X_1)+H(X_2)+H(Y)$. An error pattern is defined by specifying

$$W_1 = \{ t : \tilde{j}_1^t \neq j_1^t , \ 1 \leq t \leq m \}$$

$$W_2 = \{ t : \tilde{j}_2^t \neq j_2^t , \ 0 \leq t \leq m \}$$

Specifying W_1 and W_2 determines which of the four cases the k-th location of the r-sequences x_1^+, x_2 and y^r fall into. Knowing the error pattern W_1, W_2 and the delays D_1, D_2, we can derive K_1, K_2, K_3, K_4, the collections of k that belong to the four respective cases. Thus we shall use the notation $P(x^+, x_2, y^r/W_1W_2)$ instead of $P(x_1^{+}(\tilde{j}_1^{(m)}), x_2(\tilde{j}_2^{(m)+}), y^r/j_1^{(m)} j_2^{(m)+})$.

We define T_ϵ^r as the set of ϵ-typical r-sequences x_1^+, x_2 y^r that have $-1/r \log P(x_1^+, x_2, y^r/W_1, W_2)$ close to the mean value (given W_1, W_2) of $P(x_{1k}^+, x_{2k}, y_k^r/W_1, W_2)$, for all possible error patterns W_1, W_2. Therefore

$$T_\epsilon^r = \{ (x_1^+, x_2, y^r) \in X^r \times X^r \times Y^r$$

$$: 1/r \left| -\log P(x_1^+, x_2, y^r/W_1, W_2) - \right.$$

$$(|K_1| [H(X_1X_2Y)] + |K_2| [H(X_1)+H(X_2Y)] +$$

$$\left. |K_3| [H(X_1Y)+H(X_2)] + |K_4| [H(X_1)+H(X_2)+H(Y)]) \right| < \epsilon$$

$$\text{for all } W_1 \subseteq \{1,\ldots,m\} \text{ and } W_2 \subseteq \{0,1,\ldots,m\} \quad \}$$

where $r=(m+1)n$. Note that T_ϵ^r is a set of sequences defined in terms of typicality, and has nothing to do with the actual received sequence or each instance of decoding. That is why we use the notation $P(x_1^+, x_2, y^r/W_1, W_2)$ instead. T_ϵ^r is introduced to provide a convenient bound on $P(x_1^+, x_2, y^r/W_1, W_2)$

For the code pair $C = (C_1, C_2)$ and messages $j_1^{(m)}$ and $j_2^{(m)+}$, the sequence error probability is defined as

$$P_{e,1}(C, j_1^{(m)}, j_2^{(m)+}) = P(\hat{J}_1^{(m)} \neq j_1^{(m)} / j_1^{(m)}, j_2^{(m)+})$$

Since

$$P_{e,1}(C, j_1^{(m)}, j_2^{(m)+}) \geq \langle p_{e,1}^t \rangle = 1/nR_1 \sum_{k=1}^{n} P(\hat{J}_{1k}^t \neq j_{1k}^t)$$

for all $1 \leq t \leq m$, it is sufficient to prove the direct part by showing that it is possible to make $P_{e,1}(C, j_1^{(m)} \ldots, j_2^{(m)+} \ldots)$ vanishingly small for some codes C, for all $j_1^{(m)}$ and $j_2^{(m)+}$. Instead of computing $P_{e,1}(C, j_1^{(m)} \ldots, j_2^{(m)+} \ldots)$, we shall evaluate the more easily computable ensemble sequence error probability

$$\bar{P}_{e,1}(\mathbb{C}) = \underset{C\in\mathbb{C}}{E} \; \underset{j_1^{(m)}, j_2^{(m)+}}{E} P_{e,1}(C, j_1^{(m)}, j_2^{(m)+})$$

Due to the symmetry of the random code, the fact that $j_1^{(m)}$ and $j_2^{(m)+}$ do not contain repeated messages and that all $j_1^{(m)}$, $j_2^{(m)+}$ are equally likely, the expectation over $j_1^{(m)}$, $j_2^{(m)+}$ may be dropped since $P_{e,1}(C, j_1^{(m)}, j_2^{(m)+})$ is the same for all $j_1^{(m)}$ and $j_2^{(m)+}$. Thus

$$\bar{P}_{e,1}(\mathbb{C}) = \underset{C\in\mathbb{C}}{E} P_{e,1}(C, j_1^{(m)}, j_2^{(m)+}) = \bar{P}_{e,1}(\mathbb{C}, j_1^{(m)}, j_2^{(m)+})$$

in which $\bar{P}_{e,1}(\mathbb{C}, j_1^{(m)}, j_2^{(m)+})$ is the value $P(\hat{J}_1^{(m)} \neq j_1^{(m)} / j_1^{(m)}, j_2^{(m)+})$, with $\mathbb{C}$ treated as a source of randomness.

Over the random code and the random channel, define the events

$$F_{\tilde{j}_1^{(m)}, \tilde{j}_2^{(m)+}} = \{ (X_1^+(\tilde{j}_1^{(m)}), X_2(\tilde{j}_2^{(m)+}), Y^r) \in T_\epsilon^r \}$$

Over the ensemble of codes, the sequence error probability is then

$$\bar{P}_{e,1}(\mathbb{C}, j_1^{(m)}, j_2^{(m)+})$$

$$\leq P(\{X_1^+(j_1^{(m)}), X_2(j_2^{(m)+}), Y^r) \notin T_\epsilon^r\} \cup \left(\bigcup_{\substack{\tilde{j}_1^{(m)}, \tilde{j}_2^{(m)} \\ \tilde{j}_1^{(m)} \neq j_1^{(m)}}} F_{\tilde{j}_1^{(m)}, \tilde{j}_2^{(m)+}} / j_1^{(m)} j_2^{(m)+} \right)$$

$$\leq P(X_1^+(j_1^{(m)}), X_2(j_2^{(m)+}), Y^r) \notin T_\epsilon^r) + \sum_{\substack{\tilde{j}_1^{(m)}, \tilde{j}_2^{(m)} \\ \tilde{j}_1^{(m)} \neq j_1^{(m)}}} P(F_{\tilde{j}_1^{(m)} \tilde{j}_2^{(m)+}} / j_1^{(m)} j_2^{(m)+})$$

The first term is less than ϵ provided r is larger than some r_0, which is proved in appendix 2.2 using the law of large numbers.

Next, consider each term in the summation

$$P(F_{\tilde{j}_1^{(m)}, \tilde{j}_2^{(m)+}} / j_1^{(m)} j_2^{(m)+})$$

$$= \sum_{T_\epsilon^r} P(x_1^+ (\tilde{j}_1^{(m)}), x_2 (\tilde{j}_2^{(m)+}), y^r / j_1^{(m)} j_2^{(m)+}) \qquad a$$

$$\leq \sum_{T_\epsilon^r} \exp_2(-|K_1| H(X_1X_2Y) - |K_2| [H(X_1)+H(X_2Y)] -$$

$$|K_3| [H(X_1Y)+H(X_2)] - |K_4| [H(X_1)+H(X_2)+H(Y)] + r\epsilon) \qquad b$$

$$= |T_\epsilon^r| \exp_2(\text{" " "}) \qquad c$$

$$\leq \exp_2(r(H(X_1X_2Y)+\epsilon)) \exp_2(\text{" " "}) \qquad d$$

$$= \exp_2(-|K_2| I(X_1;X_2Y) - |K_3| I(X_2;X_1Y) - |K_4| I(X_1X_2;Y) + 2r\epsilon)) \qquad e$$

$$= \exp_2(-|K_2| I(X_1;Y/X_2) - |K_3| I(X_2;Y/X_1) - |K_4| I(X_1X_2;Y) + 2r\epsilon)) \qquad f$$

in which a follows from the definition of $F_{\tilde{j}_1^{(m)}, \tilde{j}_2^{(m)+}}$, and b follows from the definition of T_ϵ^r. The value of $|T_\epsilon^r|$ in c is overbounded in d by the following argument. Consider specifically $W_1 = W_2 = \phi$, hence $|K_1| = r$, $|K_2| = |K_3| = |K_4| = 0$. From the definition of T_ϵ^r, we have

$$P(x_1^+ (\tilde{j}_1^{(m)}), x_2 (\tilde{j}_2^{(m)+}), y^r / W_1 = W_2 = \phi)$$

$$\geq \exp_2(-r (H(X_1X_2Y) + \epsilon)) \quad .... *$$

Since

$$1 \geq \sum_{\substack{(x_1^+, x_2, y^r / W_1, W_2) \\ \text{satisfying } *}} P(x_1^+ (\tilde{j}_1^{(m)}), x_2 (\tilde{j}_2^{(m)+}), y^r / W_1 = W_2 = \phi)$$

$$\geq \sum_{T_\epsilon^r} P(x_1^+ (\tilde{j}_1^{(m)}), x_2 (\tilde{j}_2^{(m)+}), y^r / W_1 = W_2 = \phi)$$

$$\geq \sum_{T_\epsilon^r} \exp_2(-\, r(H(X_1X_2Y) + \epsilon)\,)$$

$$= |T_\epsilon^r|\ \exp_2(-\, r(H(X_1X_2Y) + \epsilon)\,)$$

As a consequence

$$|T_\epsilon^r| \leq \exp_2(\, r\ (H(X_1X_2Y) + \epsilon)\,)$$

It follows then

$$\bar{P}_{e,1}(\mathbb{C}, j_1^{(m)}, j_2^{(m)+})$$

$$\leq \epsilon + \sum_{\substack{\tilde{j}_1^{(m)}, \tilde{j}_2^{(m)+} \\ \tilde{j}_1^{(m)} \neq j_1^{(m)}}} \exp_2(-|K_2|I(X_1;Y/X_2) - |K_3|I(X_2;Y/X_1) - |K_4|I(X_1X_2;Y) + 2r\epsilon) \quad a$$

$$= \epsilon + \sum_{\substack{W_1 \neq \phi \\ W_2}} \exp_2(nR_1|W_1| + nR_2|W_2|)\cdot$$

$$\exp_2(-|K_2|I(X_1;Y/X_2) - |K_3|I(X_2;Y/X_1) - |K_4|I(X_1X_2;Y) + 2r\epsilon) \quad b$$

$$= \epsilon + \sum_{\substack{W_1 \neq \phi \\ W_2}} \exp_2(\,|K_1|\,2\epsilon + |K_2|\,(R_1 - I(X_1;Y/X_2) + 2\epsilon\,) +$$

$$|K_3|(R_2 - I(X_2;Y/X_1) + 2\epsilon) + |K_4|(R_1 + R_2 - I(X_1X_2;Y) + 2\epsilon)) \quad c$$

in which b is due to the fact that there are $\exp_2(nR_1|W_1| + nR_2|W_2|)$ different possible pairs of $\tilde{j}_1^{(m)}$ and $j_2^{(m)+}$ for a given W_1 and W_2; and c is due to the fact that

$$n|W_1| = |K_2| + |K_4|$$

$$n|W_2| = |K_3| + |K_4|$$

Let R_1 and R_2 satisfy the following inequalities for some positive $\delta > 2(m+1)\epsilon$

$$R_1 < I(X_1;Y/X_2) - 2\epsilon - \delta$$

$$R_2 < I(X_2;Y/X_1) - 2\epsilon - \delta$$

$$R_1 + R_2 < I(X_1X_2;Y) - 2\epsilon - \delta.$$

we have

$$\bar{P}_{e,1}(\mathbb{C}, j_1^{(m)}, j_2^{(m)+}) \leq \epsilon + \sum_{\substack{W_1,W_2 \\ : W_1 \neq \phi}} \exp_2(|K_1|2\epsilon - \delta(|K_2| + |K_3| + |K_4|))$$

$$\leq \epsilon + \sum_{\substack{W_1,W_2 \\ : W_1 \neq \phi}} \exp_2((m+1).n.2\epsilon - n\delta)$$

$$\leq \epsilon + 2^{2m+1} \exp_2(-n(\delta - 2(m+1)\epsilon))$$

The second inequality follows from an obvious upper bound on $/K_1/$ and an obvious lower bound on $/K_2/+/K_3/+/K_4/$. Since there are only a finite number of terms (of the order 2^{2m}) in the above sum, we have for all $\delta>0$, $\bar{P}_{e,1}(\mathbb{C}, j_1^{(m)}, j_2^{(m)+})$ goes to zero for sufficiently large n. Since

$$\bar{P}_{e,1}(\mathbb{C}) = \underset{C\in\mathbb{C}}{E} \underset{\tilde{j}_1^{(m)}, \tilde{j}_2^{(m)+}}{E} P_{e,1}(C, j_1^{(m)}, j_2^{(m)+}) = P_{e,1}(\mathbb{C}, j_1^{(m)}, j_2^{(m)+})$$

as shown previously, there exists therefore at least one code C in the code ensemble that has

$$P_{e,1}(C) = \underset{j_1^{(m)} j_2^{(m)+}}{E} P_{e,1}(C, j_1^{(m)}, j_2^{(m)+})$$

go to zero for large n. Thus all points in the open set $\bar{\mathcal{R}} = \{(R_1, R_2)\}$ such that

$R_1 < I(X_1;Y/X_2)$

$R_2 < I(X_2;Y/X_1)$

$R_1 + R_2 < I(X_1X_2;Y)$

are achievable. Hence $\tilde{\mathcal{R}} \subseteq \mathcal{R}$, completing the proof for the direct part.

2.3 The converse for the two-user asynchronous channel

For each transmission rate outside the closure of the achievable region $\bar{R}$, we seek to lower bound the error probability by a positive quantity independent of the code used, thus showing that reliable transmission is not feasible at these rates. The codebooks C_i, i=1,2 map each J_i of the 2^{NR_i} messages into the code sequence $X_i(J_i)$. We shall represent J_i as a sequence of NR_i bits $J_{i,1}J_{i,2}\ldots J_{i,NR_i}$. The code sequence X_i^N consists of N code symbols $X_{i,1}X_{i,2}\ldots X_{i,N}$.

The code sequences are shifted relatively. Without loss of generality, we assume that the code sequence X_1^N remains stationary and we shall just look at decoding for the first user. This is allowable because the transmitter-receiver pair for the first user is assumed to be synchronized, either by synchronizing both clocks beforehand or by the use of preambles. The code sequence for the second user X_2^N is shifted to the left by D symbols with respect to X_1^N, as shown in figure 2.3.1. D is assumed to be uniformly distributed for the integers between 0 and (N-n), inclusive. Thus the first n symbols of X_1^N must overlap with part of X_2^N. We denote the subsequences of these n symbols of X_1^N, X_2^N and Y^N by X_1^n, X_2^n and Y^n respectively, as shown in figure 2.3.1.

The observation of Y^n should help us to estimate part, if not all, of the information conveyed by the messages J_1 and J_2. Specifically, we are interested in the estimation of a subset of nR_i bits of the set $\{J_{i,1}, J_{i,2}, \ldots, J_{i,NR_i}\}$. We denote this

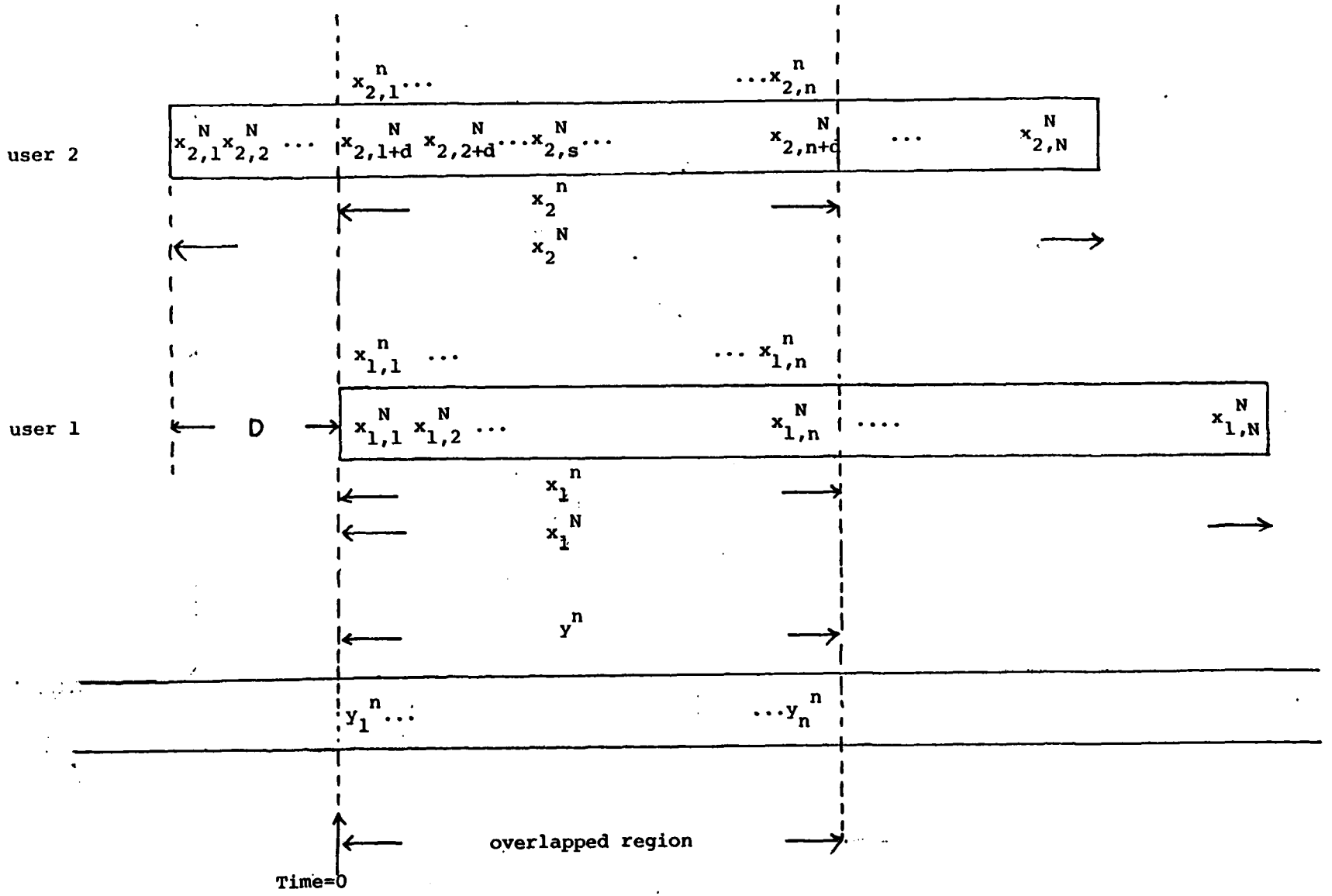


Figure 2.3.1 The overlapped region for decoding

subset by $L_i = \{L_{i,1}, L_{i,2}, \ldots, L_{i,nR_i}\}$ and its estimate from Y^n by $\hat{L}_i = \{\hat{L}_{i,1}, \hat{L}_{i,2}, \ldots, \hat{L}_{i,nR_i}\}$. Since the above model decodes only for the first user, the observation of Y^n does not provide much information about a fixed L_2 due to the uncertainty of the relative shifts. However, the statement about $\hat{L}_2$ and $<P_{e,2}>$ can be obtained by giving user 2 treatment identical to user 1. For a very asynchronous system, we shall let N go to infinity while fixing the value of n.

The converse of the coding theorem states that

For any $L_i \subseteq \{J_{i,1}, J_{i,2}, \ldots, J_{i,NR_i}\}$, the bit error probability

$$<P_{e,i}> \triangleq 1/nR_i \sum_{k=1}^{nR_i} P(\hat{L}_{i,k} \neq L_{i,k})$$

is lower bounded from zero if (R_1, R_2) is outside $\bar{R}$.

The following arguments relate the entropy of L_i given the channel output sequence with the bit error probability of the estimates through a form of Fano's inequality.

$$H(L_1/Y^n L_2)$$

$$= H(L_1/Y^n \hat{L}_1 L_2) \qquad 2.3.1a$$

$$\leq H(L_1/\hat{L}_1) \qquad b$$

$$\leq \sum_{k=1}^{nR_1} H(L_{1,k}/\hat{L}_1) \qquad c$$

$$\leq \sum_{k=1}^{nR_1} H(L_{1,k}/\hat{L}_{1,k}) \qquad d$$

$$\leq \sum_{k=1}^{nR_1} H_e(P(\hat{L}_{1,k} \neq L_{1,k})) \; ; \; H_e(x) = -x \log x - (1-x) \log (1-x) \qquad e$$

$$\leq n R_1 H_e\left(1/nR_1 \sum_{k=1}^{nR_1} P(\hat{L}_{1,k} \neq L_{1,k}) \right) \qquad f$$

$$= n R_1 H_e(\langle P_{e,1} \rangle) \qquad g$$

in which a results because L_1, Y^n, $\hat{L}_1$ is a Markov chain (figure 2.3.2); b follows because entropy cannot be decreased by unconditioning; c is due to the fact that the entropy of a set of random variables is less than the sum of the entropies of the random variables; d is due to the fact that entropy cannot be decreased by unconditioning; e is due to the usual Fano's inequality (equation 4.3.4 with M=2 in [9]); f is due to the fact that H_e is a convex $\cap$ function; and g introduces a new notation. Thus we have

Theorem 2.3.1 (Fano's inequality)

$$H(L_1/Y^n L_2) \leq n R_1 H_e(\langle P_{e,1} \rangle) \qquad 2.3.2\ a$$

$$H(L_2/Y^n L_1) \leq n R_2 H_e(\langle P_{e,2} \rangle) \qquad b$$

$$H(L_1 L_2/Y^n) \leq n(R_1+R_2) H_e(R_1/(R_1+R_2) \langle P_{e,1} \rangle + R_2/(R_1+R_2) \langle P_{e,2} \rangle) \qquad c$$

in which b is derived similarly as a. The proof of c is slightly different, for which

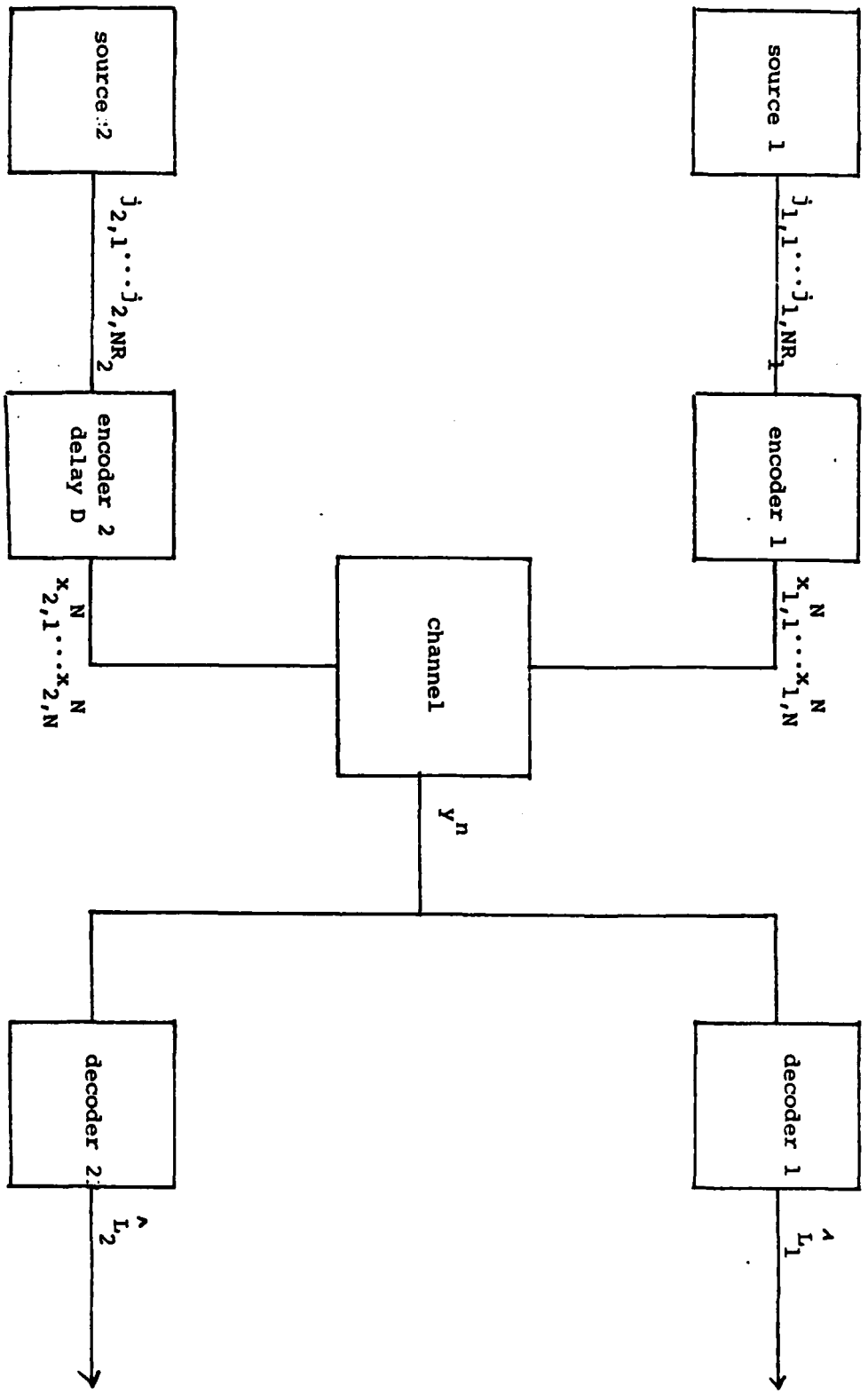


Figure 2.3.2 The Markov chain process of encoding-decoding.

$$H(L_1 L_2 / Y^n)$$

$$= H(L_1 L_2 / Y^n \hat{L}_1 \hat{L}_2) \qquad 2.3.3a$$

$$\leq H(L_1 L_2 / \hat{L}_1 \hat{L}_2) \qquad b$$

$$\leq \sum_{k=1}^{nR_1} H(L_{1,k} / \hat{L}_1 \hat{L}_2) + \sum_{k=1}^{nR_2} H(L_{2,k} / \hat{L}_1 \hat{L}_2) \qquad c$$

$$\leq \sum_{k=1}^{nR_1} H(L_{1,k} / \hat{L}_{1,k}) + \sum_{k=1}^{nR_2} H(L_{2,k} / \hat{L}_{2,k}) \qquad d$$

$$\leq \sum_{k=1}^{nR_1} H_e(P(\hat{L}_{1,k} \neq L_{1,k})) + \sum_{k=1}^{nR_2} H_e(P(\hat{L}_{2,k} \neq L_{2,k})) \qquad e$$

$$\leq nR_1 H_e(1/nR_1 \sum_{k=1}^{nR_1} P(\hat{L}_{1,k} \neq L_{1,k})) + nR_2 H_e(1/nR_2 \sum_{k=1}^{nR_2} P(\hat{L}_{2,k} \neq L_{2,k})) \qquad f$$

$$= nR_1 H_e(\langle P_{e,1} \rangle) + nR_2 H_e(\langle P_{e,2} \rangle) \qquad g$$

$$\leq n(R_1+R_2) H_e(R_1/(R_1+R_2) \langle P_{e,1} \rangle + R_2/(R_1+R_2) \langle P_{e,2} \rangle) \qquad h$$

in which the line of reasoning from a to g is the same as that of (2.3.1 a-g); and h follows from the fact that $H_e(x)$ is a convex $\cap$ function.

Q.E.D.

Furthermore

$$H(L_1/L_2) = n R_1 \qquad 2.3.4 a$$

$$H(L_2/L_1) = n R_2 \qquad b$$

$$H(L_1 L_2) = n(R_1 + R_2) \qquad c$$

Subtracting (2.3.2) from (2.3.4) gives

$$n R_1 H_e(\langle P_{e,1} \rangle) \geq n R_1 - I(L_1;Y^n/L_2) \qquad 2.3.5\ a$$

$$n R_2 H_e(\langle P_{e,2} \rangle) \geq n R_2 - I(L_2;Y^n/L_1) \qquad b$$

$$n(R_1+R_2)\, H_e(R_1/(R_1+R_2)\, \langle P_{e,1} \rangle + R_2/(R_1+R_2)\, \langle P_{e,2} \rangle)$$

$$\geq n(R_1 + R_2) - I(L_1 L_2;Y^n) \qquad c$$

The following arguments are mostly concerned with upper bounding the values of the mutual information in the above inequalities.

Theorem 2.3.2 (Data processing theorem)

$$I(L_1;Y^n/L_2) \leq I(X_1^n;Y^n/X_2^n) \qquad 2.3.6\ a$$

$$I(L_2;Y^n/L_1) \leq I(X_2^n;Y^n/X_1^n) \qquad b$$

$$I(L_1 L_2;Y^n) \leq I(X_1^n X_2^n;Y^n) \qquad c$$

Proof

Consider

$$I(X_1^n L_1;Y^n/L_2 X_2^n)$$

$$= I(X_1^n;Y^n/L_2 X_2^n) + I(L_1;Y^n/L_2 X_1^n X_2^n) \qquad 2.3.7$$

The second term equals zero since Y^n provides no new information, given X_1^n, about L_1, due to the fact that L_1, X_1^n, Y^n is a Markov chain. The first term equals $I(X_1^n;Y^n/X_2^n)$ since L_2, X_2^n, Y^n is a Markov chain. Hence

$$I(X_1^n L_1; Y^n / L_2 X_2^n)$$

$$= I(X_1^n; Y^n / X_2^n) \qquad 2.3.8$$

On the other hand,

$$I(X_1^n L_1; Y^n / L_2 X_2^n)$$

$$= I(L_1; Y^n / L_2 X_2^n) + I(X_1^n; Y^n / L_1 L_2 X_2^n) \qquad 2.3.9$$

in which the first term equals

$$I(L_1; Y^n L_2 X_2^n) \qquad 2.3.10\ a$$

$$\geq I(L_1; Y^n L_2) \qquad b$$

$$= I(L_1; Y^n / L_2) \qquad c$$

with a due to the fact that for A independent of C

$$I(A; BC)$$

$$= I(A; C) + I(A; B/C)$$

$$= I(A; B/C) \qquad 2.3.11$$

b is due to the fact that mutual information is never increased by providing less observation and c follows from the identity (2.3.11). The second term in (2.3.9) is trivially lower bounded by zero. Hence putting (2.3.10) into (2.3.9) gives

$$I(L_1; Y^n / L_2)$$

$$\leq I(X_1^n L_1; Y^n / L_2 X_2^n) \qquad 2.3.12$$

Combining (2.3.8) and (2.3.12) gives

$$I(L_1;Y^n/L_2)$$

$$\leq I(X_1^n;Y^n/X_2^n)$$

which is (2.3.6a). The proofs for (2.3.6b and c) follow the same argument.

Q.E.D.

Let the subscript k denote the k-th symbol of a sequence. The following theorem upper bounds $I(X_1^n ; Y^n/X_2^n)$ in terms of the statistics of the codebooks. Recall that $X_{1,s}^N$, $X_{2,s}^N$, $1\leq s\leq N$, are the random variables for the s-th symbol for the codewords in the codebooks C_1 and C_2 respectively.

Theorem 2.3.3

$$I(X_1^n;Y^n/X_2^n) \leq \sum_{k=1}^{n} I(X_{1,k}^n ;Y_k^n /X_{2,k}^*) \qquad 2.3.13\ a$$

$$I(X_2^n ;Y^n/X_1^n) \leq \sum_{k=1}^{n} I(X_{2,k}^* ;Y_k^n /X_{1,k}^n) \qquad b$$

$$I(X_1^n X_2^n ;Y^n) \leq \sum_{k=1}^{n} I(X_{1,k}^n X_{2,k}^* ;Y_k^n) \qquad c$$

in which

$$P(X_{1,k}^n) = P(X_{1,k}^N) \qquad d$$

$$P(X_{2,k}^*) = 1/(N-n+1) \sum_{s=k}^{N-n+k} P(X_{2,s}^N) \qquad e$$

Remark:

The statistics of $X_{1,k}^{n}$ is the same as $X_{1,k}^{N}$ since we treat X_1^N as unshifted in time. Consider the statistics of $X_{2,k}^{n}$. Since the delay is evenly distributed in the region 0 to N-n, only the s-th symbol with $k \leq s \leq N-n+k$ can fall into the k-th position, with probability $1/(N-n+1)$. Hence $X_{2,k}^{n}$ is a random variable with statistics that equals the average statistics of the code symbols that can possibly fall into the k-th position. Thus from the point of view of the first user at the k-th position, the second user has statistics equal to that of $X_{2,k}^{*}$ defined in (2.3.13e).

Proof:

$$I(X_1^n ; Y^n / X_2^n)$$

$$= \sum_{k=1}^{n} I(Y_k^n ; X_1^n / X_2^n Y_1^n Y_2^n \ldots Y_{k-1}^n) \qquad 2.3.14\ a$$

$$= \sum_{k=1}^{n} [H(Y_k^n / X_2^n Y_1^n Y_2^n \ldots Y_{k-1}^n) - H(Y_k^n / X^n, X_2^n Y_1^n Y_2^n \ldots Y_{k-1}^n)] \qquad b$$

$$\leq \sum_{k=1}^{n} [H(Y_k^n / X_{2,k}^n) - H(Y_k^n / X_{1,k}^n X_{2,k}^n)] \qquad c$$

$$= \sum_{k=1}^{n} I(X_{1,k}^n ; Y_k^n / X_{2,k}^n) \qquad d$$

in which a, b and d are due to well-known information theoretic identities. The first term in the square bracket in b is upper bounded by $H(Y_k^n / X_{2,k}^n)$ since unconditioning cannot increase the entropy. The second term in b equals $H(Y_k^n / X_{1,k}^n X_{2,k}^n)$ since the channel is memoryless. The equations (2.3.13 b and c) are proved similarly. Furthermore, the joint probability distribution for $I(X_{1,k}^n ; Y_k^n / X_{2,k}^n)$ is

$$= P(X_{1,k}^{n} X_{2,k}^{n} Y_k^{n})$$

$$= \sum_d P(d)\, P(X_{1,k}^{n} X_{2,k}^{n} Y_k^{n} / d) \qquad 2.3.15a$$

$$= \sum_d P(d)\, P(X_{1,k}^{N} X_{2,k+d}^{N} Y_k^{n}) \qquad b$$

$$= \sum_d P(d)\, P(X_{1,k}^{N} X_{2,k+d}^{N})\, P(Y_k^{n} / X_{1,k}^{N} X_{2,k+d}^{N}) \qquad c$$

$$= \sum_d P(d)\, P(X_{1,k}^{N})\, P(X_{2,k+d}^{N})\, P(Y_k^{n} / X_{1,k}^{N} X_{2,k+d}^{N}) \qquad d$$

$$= P(X_{1,k}^{N})\, P(Y_k^{N} / X_{1,k}^{N} X_{2,k+d}^{N}) \sum_d P(d)\, P(X_{2,k+d}^{N}) \qquad e$$

$$= P(X_{1,k}^{N})\, P(X_{2,k}^{*})\, P(Y_k^{N} / X_{1,k}^{N} X_{2,k}^{*}) \qquad f$$

where

$$P(X_{2,k}^{*}) = \sum_d P(d)\, P(X_{2,k+d}^{N})$$

in which a and c follows from conditioning; b follows from the fact that $x_{1,k}^{n} = x_{1,k}^{N}$, and $x_{2,k}^{n} = x_{2,k+d}^{N}$ for given d; d follows from the independence of $X_{1,k}^{N}$ and $X_{2,k+d}^{N}$; e follows from the fact that the channel transition probability is independent of d given $x_{1,k}^{N}, x_{2,k+d}^{N}$; and f defines the random variable $X_{2,k}^{*}$.

Q.E.D.

If we make N large while keeping n constant, it follows readily from (2.3.13 e) that $P(X_{2,k}^{*})$ are almost identical for all $1 \leq k \leq n$ since the boundary effect may be ignored. Thus we shall

approximate $P(X^*_{2,k})$ by a limiting distributions $P_2(X_2)$. Subsequently, (2.3.13 a, b and c) become

$$I(X_1^n ; Y^n / X_2^n) \leq \sum_{k=1}^{n} I(X_{1,k}^N ; Y_k^n / X_2) \qquad 2.3.16\ a$$

$$I(X_2^n ; Y^n / X_1^n) \leq \sum_{k=1}^{n} I(X_2 ; Y_k^n / X_{1,k}^N) \qquad b$$

$$I(X_1^n X_2^n ; Y^n) \leq \sum_{k=1}^{n} I(X_{1,k}^N X_2 ; Y_k^n) \qquad c$$

Since mutual information is a convex $\cap$ function of the input probability distributions $X_{1,k}^N$., the right hand side of (2.3.16 a) is further upper bounded by $n.I(X_1;Y_{\cdot}/X_2)$ in which

$$P_1(X_1) = 1/n \sum_{k=1}^{n} P(X_{1,k}^N)$$

and

$$P(x_1 x_2 y) = P_1(x_1)\, P_2(x_2)\, P(y/x_1 x_2)$$

Similarly, the right hand side of (2.3.16 c) is upper bounded by $n.I(X_1X_2;Y)$. The right hand side of (2.3.16 b) equals $n.I(X_2;Y/X_1)$ due to linearity in P(d) for mutual information conditioned on d. We omit a formal proof for the inequality

$$I(X_2;Y_k^n/X_{1,k}^N) \leq I(X_2;Y/X_1)$$

because this inequality can be obtained when we consider decoding for user 2, using the same reasoning as that leading to

$$I(X_{1,k}^N ; Y_k^n / X_1) \leq I(X_1 ; Y/X_2)$$

when we consider decoding for user 1.

Combining the chain of inequalities gives

$$n R_1 \quad H_e(\langle P_{e,1} \rangle) \geq n R_1 - n I(X_1;Y/X_2) \qquad 2.3.17\ a$$

$$n R_2 \quad H_e(\langle P_{e,2} \rangle) \geq n R_2 - n I(X_2;Y/X_1) \qquad b$$

$$n(R_1+R_2)\ H_e(R_1/(R_1+R_2)\ \langle P_{e,1} \rangle + R_2/(R_1+R_2)\ \langle P_{e,2} \rangle)$$

$$\geq n(R_1+R_2) - n I(X_1X_2;Y) \qquad c$$

Since $H_e(x)$ is a nondecreasing function in the interval $[0,1/2]$, we may express finally the converse of the coding theorem for the two-user asynchronous multiple access channel in the following form

Theorem 2.3.4

$$\langle P_{e,1} \rangle \geq H_e^{-1}(1-I(X_1;Y/X_2)/R_1) \qquad 2.3.18\ a$$

$$\langle P_{e,2} \rangle \geq H_e^{-1}(1-I(X_2;Y/X_1)/R_2) \qquad b$$

$$R_1/(R_1+R_2)\ \langle P_{e,1} \rangle + R_2/(R_1+R_2)\ \langle P_{e,2} \rangle$$

$$\geq H_e^{-1}(1-I(X_1X_2;Y)/(R_1+R_2) \qquad c$$

in which for $0 \leq y \leq 1$, the function $H_e^{-1}(y)=x$ such that $H_e(x)=y$; and for $y<0$, $H_e^{-1}(y)=0$. Hence reliable communication is not feasible if

$$R_1 > I(X_1;Y/X_2)$$

$R_2 > I(X_2;Y/X_1)$

or

$R_1 + R_2 > I(X_1X_2;Y)$

for all mutually independent probability distributions $P_1(X_1)$ and $P_2(X_2)$. Therefore, $\mathcal{R} \subset \bar{\mathcal{R}}$.

Theorem 2.3.4 concludes the proof of the converse.

Appendix 2.1 Preambles for synchronization

Section 2.2 assumes that the preambles can be detected and correctly located in time. This section shall give a description of this synchronization mechanism and compute the probability of failure to achieve synchronization, averaged over the code ensemble defined in section 2.2.

The synchronization mechanism works as follows. Suppose the preamble sequence b_n for the first user starts at the s_0 -th symbol of channel output sequence Y^{mn}. (Recall that m is the number of codewords between two preambles, and n is the length of a codeword.) Let $Y_{(s)}$ be the subsequence of Y^{mn} which starts at the s-th symbol and ends at the (s+n-1)-th symbol of Y^{mn}. Let

$$S = \{s : (b_n, y_{(s)}) \in T_\epsilon^n \}$$

in which T_ϵ^n is the set of ϵ-typical n-sequences defined by

$$T_\epsilon^n = \{ (x_1^n, y^n) \in X_1^n \times Y^n$$

$$: 1/n \left|-\log P(x_1^n, y^n) - H(X_1,Y)\right| < \epsilon$$

$$1/n \left|-\log P(x_1^n) - \log P(y^n) - H(X_1) - H(Y)\right| < \epsilon \}$$

Notice that we do not require joint typicality with the codewords of the second user while we obtain the synchronization of the first user. The effect of the code sequence of the second user is included in P(Y), which depends on $P_2(X_2)$.

Synchronization failure for a given code C generated in random is defined as the event

$$F(C) = (s_0 \notin S) \cup (\text{some } s \neq s_0 \text{ and } s \in S)$$

Over the set of random codes $\mathbb{C}$, the event of synchronization failure is given by

$$F(\mathbb{C}) = \left((B_n, Y_{(s_0)}) \notin T_\epsilon^n \right) \cup \left(\bigcup_{s \neq s_0} \left((B_n, Y_{(s)}) \in T_\epsilon^n \right) \right)$$

Hence using the union bound

$$P(F(\mathbb{C})) \leq P\left((B_n, Y_{(s_0)}) \notin T_\epsilon^n \right) + \sum_{\substack{s=1 \\ s \neq s_0}}^{m^n} P\left((B_n, Y_{(s)}) \in T_\epsilon^n \right)$$

The first term is upper bounded by ϵ, by the definition of ϵ-typical sequences provided that n is larger than some n_σ. Let the subscript k denote the k-th symbol of a sequence and consider each term in the above summation.

$$P\left((B_n, Y_{(s)}) \in T_\epsilon^n \right)$$

$$= \sum_{T_\epsilon^n} P(b_n, y_{(s)}) \qquad a$$

$$= \sum_{T_\epsilon^n} \prod_{k=1}^{n} P(b_{nk}, y_{(s)k}) \qquad b$$

$$= \sum_{T_\epsilon^n} \prod_{k=1}^{n} P(b_{nk}) \cdot P(y_{(s)k}) \qquad c$$

$$\leq \sum_{T_\epsilon^n} \exp_2\left(-n \left(H(X_1) + H(Y) - \epsilon \right) \right) \qquad d$$

$$\leq \exp_2\left(n(H(X_1 Y) + \epsilon) \right) \cdot \exp_2\left(-n \left(H(X_1) + H(Y) - \epsilon \right) \right) \qquad e$$

$$= \exp_2\left(-n \left(I(X_1;Y) - 2\epsilon \right) \right) \qquad f$$

in which b is due to the facts that the codewords are generated independently symbol by symbol and that the channel is memoryless; c is due to the facts that $s \neq s_0$ and that the

codewords are generated independently; the upper bound in d results from the definition of T_ϵ^n ; e is due to an upper bound on the cardinality of an ϵ-typical set, derived similarly to that in section 2.2. Hence

$$P(F(\mathbb{C}))$$

$$\leq \epsilon + \sum_{s=1, s\neq s_0}^{mn} \exp_2(-n(I(X_1;Y) - 2\epsilon))$$

$$\leq \epsilon + mn \exp_2(-n(I(X_1;Y) - 2\epsilon))$$

Thus provided

$$I(X_1;Y) - 1/n \log (mn) - 2\epsilon \geq \delta$$

or

$$1/n \exp_2(n(I(X_1;Y) - \delta - 2\epsilon)) \geq m$$

for some fixed positive δ ,

$$P(F(\mathbb{C})) \leq 2^{-n\delta}$$

thus synchronization can be made reliable.

The reliability of synchronization for the second user can be similarly established.

Appendix 2.2 On sets of ϵ-typical sequences

We want to show that for each $\epsilon > 0$, there exists an r_o such that for $r > r_o$,

$$P\left((X_1^+(j_1^{(m)}), X_2(J_2^{(m)+}), Y^r) \notin T_\epsilon^r \right) \leq \epsilon$$

in which

$$T_\epsilon^r = \{ (x_1^+, x_2, y^r)$$

$$: 1/r \left| -\log P(x_1^+, x_2, y^r / W_1, W_2) - \right.$$

$$\left(|K_1| [H(X_1 X_2 Y)] + |K_2| [H(X_1)+H(X_2 Y)] + \right.$$

$$\left. |K_3| [H(X_1 Y)+H(X_2)] + |K_4| [H(X_1)+H(X_2)+H(Y)] \right) | < \epsilon$$

$$\text{for all } W_1 \subseteq \{1,\ldots,m\},\ W_2 \subseteq \{0,1,\ldots,m\} \}$$

Proof:

Consider the random variables

$$z_k(W_1, W_2) \triangleq -\log P(x_{1k}^+, x_{2k}, y_k^r / W_1, W_2)$$

Let $h(k) = i$ if $k \in K_i$. Hence $E[z_k(W_1, W_2)] = H_{h(k)}$ where

$$H_i = H(X_1 X_2 Y) \quad \text{if} \quad i = 1$$

$$= H(X_1) + H(X_2 Y) \quad i = 2$$

$$= H(X_1 Y) + H(X_2) \quad i = 3$$

$$= H(X_1) + H(X_2) + H(Y) \quad i = 4$$

Define the random variable

$$u_r(W_1,W_2) = 1/r \sum_{k=1}^{r} z_k(W_1,W_2)$$

Thus, we may equivalently express T_ϵ^n as

$$T_\epsilon^n = \{(x_1^r,x_2,y^r) \epsilon X_1^r X X_2^r X Y^r : |u_r(W_1,W_2) - \bar{u}_r(W_1,W_2)| < \epsilon \text{ for all } W_1,W_2\}$$

By the Chebychev inequality,

$$P(|u_r(W_1,W_2) - \bar{u}_r(W_1,W_2) \geq \epsilon)$$

$$\leq \sigma^2_{u_r(W_1,W_2)} / \epsilon^2$$

$$= 1/r^2 \sum_{k=1}^{r} \sigma^2_{z_k(W_1,W_2)} / \epsilon^2$$

$$\leq 1/r^2 \sum_{k=1}^{r} \sigma^2_{max} / \epsilon^2$$

$$= \sigma^2_{max} / r \epsilon^2$$

where σ^2_{max} is the maximum value of the four possible values of the variances $\sigma^2_{z_k(W_1,W_2)}$. In order that $(x_1^r,x_2,y^r) \notin T_\epsilon^r$, it is necessary that $|u_r(W_1,W_2)-u_r(W_1,W_2)| \geq \epsilon$ for some W_1,W_2, which occurs with a probability upper bounded (through union bounding) by

$$\sum_{W_1,W_2} P(|u_r(W_1,W_2) - \bar{u}_r(W_1,W_2)| \geq \epsilon)$$

$$\leq \sum_{W_1,W_2} \sigma^2_{max} / r \epsilon^2$$

$$= 2^{2nR} \sigma^2_{max} / r \epsilon^2$$

Hence

$$P\left((X_1^r, X_2, Y^r) \notin T_\epsilon^r \right)$$

$$\leq 2^{2m+1} \sigma_{max}^2 / r \epsilon^2$$

$$\leq \epsilon$$

if

$$r > r_0 = [\, 2^{2m+1} \sigma_{max}^2 / \epsilon^3 \,]$$

Q.E.D.

Chapter 3 The Multiple Access Channel with a Variable Set of Simultaneous Users.

3.1 The coding theorem

Consider a communication system with M synchronous users, each generating n-bit packets at a Poisson rate of λ/M. Each packet must be received with a delay of less than D from its generation time. For channels with a large bandwidth which is used by a large number of users, the maximum tolerable delay can be a significant factor in the design of the access scheme. As shown in section 1.3, the round-robin TDMA scheme would incur an average delay of $M/[2(1-\lambda)]$ slots. Thus for large M or λ close to one, the delay becomes intolerable. We are interested in the case when the *maximum* tolerable delay D is much smaller than $M/[2(1-\lambda)]$.

This chapter shows that imposing the bound D on delay reduces the maximum throughput of the channel. An intuitive explanation goes as follows. Within the delay D, the number of active users (those with a message to send) is $N = D\lambda + O(\sqrt{D\lambda})$. We assume that $N << M$. Thus the N active users would have to contend for the D ($<< M$) slots. Without feedback, conflict free scheduling is impossible. To ensure reliable communication within the delay D, it seems that an active user should transmit the coded message as soon as it is generated, and spread the message over a time period of D. Such a system behaves asynchronously, even though the transmitters are synchronized.

Theorem 3.1.1 generalizes and makes precise this intuitive notion of "asynchronism" as a result of the uncertainty about the set of active users. The theorem is based on the model that exactly N out of M synchronous users (N << M) are active, and that each of the M choose N combinations are equally likely. We assume that there is a zero element in the code alphabet X (which is the same for all users) so that an inactive user would transmit the zero element. The codebook C_i of user i consists of 2^{nR_i} codewords. The effective rate of each user is $s_i = (N/M)\, R_i$ since a user is active with probability N/M.

We assume that the channel is memoryless and symmetrical with respect to the M users, that is, $P(y/x_1 x_2 \ldots x_M)$ is the same if the input symbols for the users are interchanged. Furthermore, we assume that the users are symmetrical in what they can do.

The communication system is reliable if and only if reliable communication is achievable for any set of N users. The capacity region S is the set of $(s_1, s_2, \ldots, s_M)$ such that reliable communication is achievable for the system. The sum-throughput of the channel is defined by

$$s = s_1 + s_2 + \ldots + s_M$$

and the sum-capacity is defined by

$$s^{\#} = \max_{S} s$$

The coding theorem can be stated as follows.

Theorem 3.1.1

$$s^* = \max_{P(X)} I(X^N; Y)$$

in which X^N denotes a set of N independent and identically distributed random variables with probability distribution P(X) and Y has conditional statistics $P(y/x_1 \ldots x_M)$ where $x_i \in X$, $1 \leq i \leq N$, are the symbols put on the channel by the N active users.

As an illustration, consider the collision channel. Let $P(x) = p$ if x is an idle and $P(x) = (1-p)/2^n$ if x is a packet. It can be readily shown that for large n,

$$I(X^N; Y) = N(1-p)\, p^{N-1} \quad \text{n-bit}$$

which has a maximum value of $(1-1/N)^{N-1}$ when $p = 1 - 1/N$. Therefore $s^* = e^{-1}$ for large N. This is the same as the capacity of the asynchronous multiple access channel with a large number of users.

Before proving theorem 3.1.1, we shall prove the following coding theorem on the capacity region of the M choose N channel.

Theorem 3.1.2

Let ψ be the set of simultaneous users. (Thus $\psi \subseteq \{1, \ldots, M\}$ and $/\psi/ = N$). The capacity region S is given by

$$S = \text{convex hull} \bigcup_{P_i(X_i),\ 1\le i\le M} \tilde{S}$$

in which $(s_1, s_2, \ldots, s_M) \in \tilde{S}$ (each $\tilde{S}$ is defined by a set of $P_i(X_i)$) if

$$\sum_{i\in\Omega} s_i < N/M \quad I(\{X_i\}_{i\in\Omega}; Y/\{X_i\}_{i\in\bar{\Omega}})$$

for all $\Omega \subseteq \psi, \bar{\Omega} = \psi - \Omega$; and for all ψ.

Proof

As an illustration, figure 3.1.1 gives $\tilde{S}$ for the case of M=3 and N=2. $\tilde{S}$ is the intersection of three pentagonal cylinders (each due to a ψ). Figure 3.1.2 shows the familiar case of M=N=3.

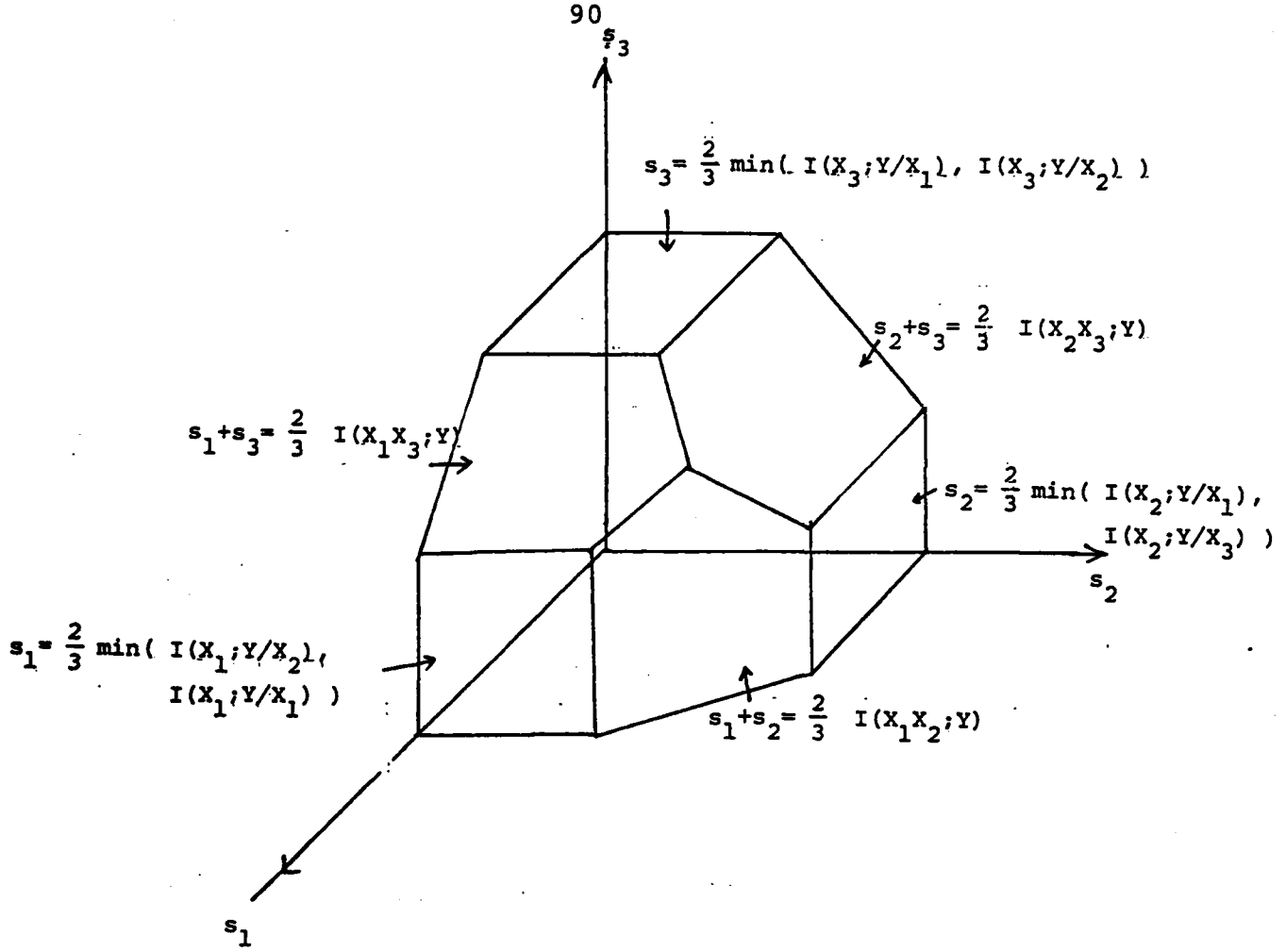
For given ψ, it is well known [16] that the capacity region for the synchronous N-user multiple access channel is given by $\mathcal{R}$ such that

$$\mathcal{R} = \text{convex hull} \bigcup_{\text{independent } P_i(X_i),\ i\in\psi} \tilde{\mathcal{R}}$$

such that $(R_1, R_2, \ldots, R_N) \in \tilde{\mathcal{R}}$ if

$$\sum_{i\in\Omega} R_i < I(\{X_i\}_{i\in\Omega}; Y/\{X_i\}_{i\in\bar{\Omega}})$$

for all $\Omega \subseteq \psi, \Omega \neq \phi$ and $\bar{\Omega} = \psi - \Omega$. Each of these $\mathcal{R}$, extended in the M dimensional space $s_1 X s_2 X \ldots X s_M$ is a cylinder. Since we require reliable communication for all ψ, the capacity region for the M choose N channel is the intersection of these cylinders, with the substitution $s_i = N/M\ R_i$. In other words,



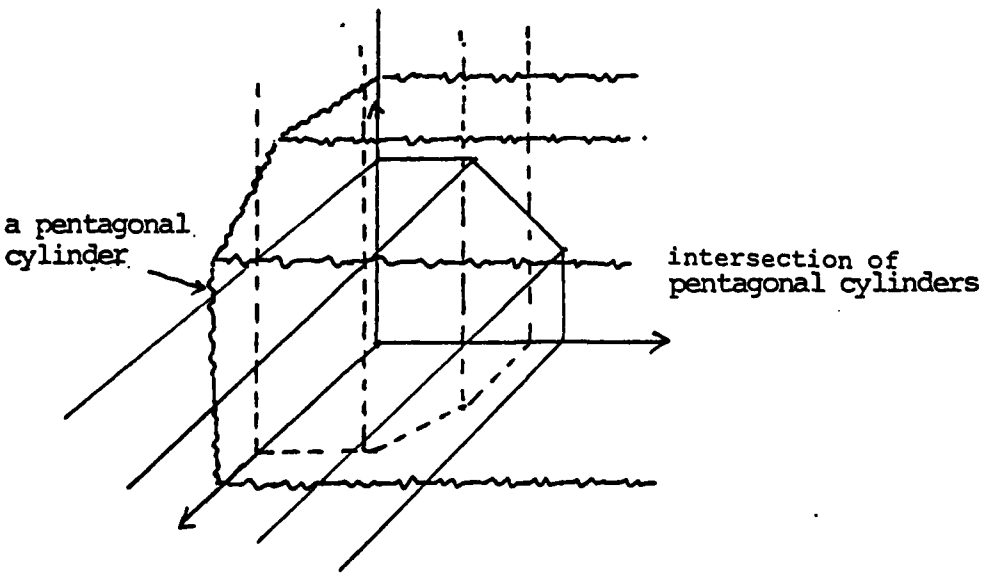


Figure 3.1.1 $\tilde{S}$ for M=3 N=2

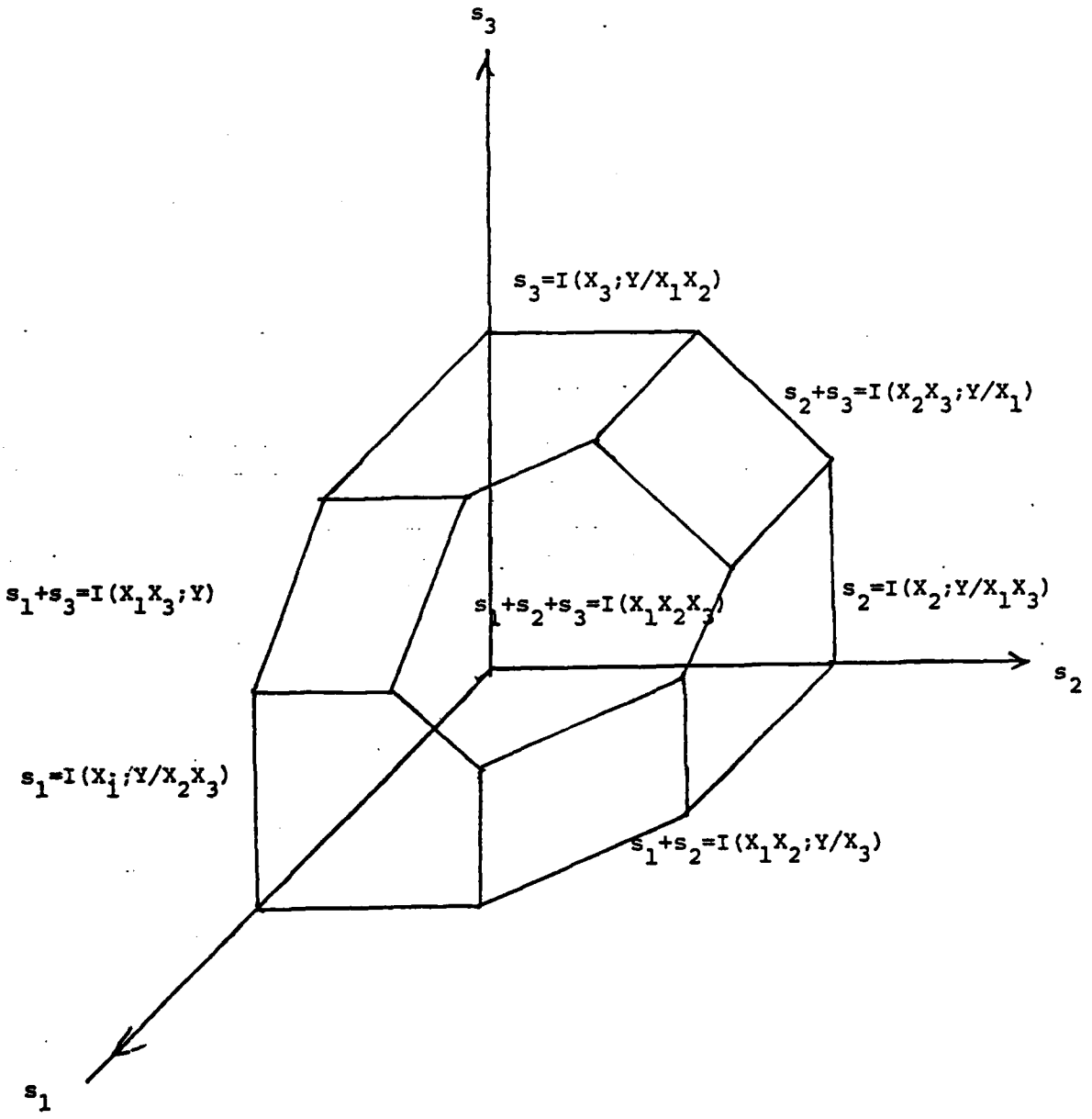


Figure 3.1.2 $\tilde{S}$ for M=N=3

S = convex hull $\bigcup_{\text{independent } P_i(X_i),\ i \in \Psi} \tilde{S}$

such that $(s_1, s_2, \ldots, s_M) \in \tilde{S}$ if

$$\sum_{i \in \Omega} s_i < N/M \quad I(\{X_i\}_{i \in \Omega}; Y / \{X_i\}_{i \in \bar{\Omega}})$$

for all $\Omega \subseteq \Psi, \Omega \neq \phi$, and $\bar{\Omega} = \Psi - \Omega$; and for all Ψ. There remains the question of identifying Ψ. For the proof for achievability, it is necessary to show that Ψ can be identified at a small cost of channel capacity. For the purpose of identification, each active source transmits a preamble sequence of length r before transmitting his codeword. The preambles are generated by random coding according to the probability distribution $P_\alpha(X)$. Appendix 3.1 describes the identification mechanism and establishes the reliability of identification of the set of active users (Ψ) in terms of the number of users M and the length of the preamble r. It is shown that the length of the preamble required is proportional to log M, which we assume to be much smaller than n, the length of the codeword. For the converse of the coding theorem, we assume the help of a genie that reveals the set of active users Ψ.

Q.E.D.

As a consequence of the above characterization of S, we have

Theorem 3.1.3

The capacity region S is convex and symmetrical in the s_i's. Thus the sum-capacity occurs along the main diagonal where the rates of the M users are all equal.

Proof

The convexity of S is established by time sharing which is permissible since the transmitters are synchronized. The region S is symmetrical in the s_i's due to the assumption of symmetry of the channel with respect to the various users, and the symmetry of what the users can do.

Q.E.D.

3.2 Proof of the direct part

Before proving the direct part, we would like to show how the model of theorem 3.1.1 applies for channels with Poisson message arrivals. Consider the following access scheme. Time is divided into periods of length D/2. All packets generated in one period are to be transmitted in the next period. Thus all messages are delivered within a delay D. The number of messages N generated in a period of D/2 averages $\lambda D/2$. For large D, N can be approximated by its average by the law of large number. Furthermore, it is unlikely that a single user would have more than one packet generated in a period, since $D << M/[2(1-\lambda)]$. If a user has more than one packet, it would treat the packets as if they are from separate sources. Thus this scheme can be approximated by the M choose N model of theorem 3.1.1. Each active source transmits a preamble sequence of length r at the beginning of a period. The message is transmitted during the rest of the period, which is of length D/2 - r symbols.

Consider letting the X_i's in theorem 3.1.2 be M independent and identically distributed random variables. Since we assume that the channel is symmetrical with respect to the M users, the quantity $I(\{X_i\}_{i\in\Omega}; Y/\{X_i\}_{i\in\bar{\Omega}})$ in theorem 3.1.2 may be abbreviated for all Ω as $I(X^K; Y/X^{N-K})$ where $K=/\Omega/$ and $N=/\Psi/$. We want to show that the point with $s_i = 1/M\ I(X^N; Y)$, for all $1\leq i\leq M$, is in the $\tilde{S}$ defined in theorem 3.1.2 with $P_i(X_i)=P(X)$ for all i. This is true if

$$\sum_{i \in \Omega} s_i$$

$$= K \left(1/M \, I(X^N ; Y)\right)$$

$$\leq N/M \, I(X^K ; Y / X^{N-K})$$

for all K such that $1 \leq K \leq N$. Equivalently, we want to show that

$$K/N \, I(X^K;Y) \leq I(X^K;Y/X^{N-K}) = I(X^N;Y) - I(X^{N-K};Y)$$

or

$$I(X^{N-K};Y) \leq (N-K)/N \quad I(X^N;Y) \qquad *$$

for all $1 \leq K \leq N$. Replacing N-K by K, * is equivalent to

$$I(X^K;Y) \leq K/N \, I(X^N;Y)$$

for all $1 \leq K \leq N$.

Proof

Consider

$$D_K = I(X^{K+1};Y) - I(X^K;Y)$$

$$= I(X;Y/X^K)$$

$$= H(X/X^K) - H(X/YX^K)$$

$$= H(X) - H(X/YX^K)$$

Due to the fact that conditioning decreases the entropy, $H(X/YX^K)$ decreases as K increases. Therefore $\{D_K\}$ is an increasing

sequence in K. Thus

$$I(X^K;Y) \leq K/N\ I(X^N;Y)$$

Q.E.D.

We have shown that the point with $s_i = 1/M\ I(X^N;\ Y)$, $1 \leq i \leq M$, is achievable. Thus an achievable sum-throughput is

$$s = \sum_{i=1}^{M} 1/M\ I(X^N;\ Y) = I(X^N;\ Y)$$

This completes the proof of the direct part of theorem 3.1.1.

3.3 Proof of the converse

This section proves that reliable communication is not possible if the sum-throughput s is above the sum-capacity

$$s^* = \max_{P(X)} I(X^N ; Y)$$

for M >> N. We shall also show that for the collision channel with $N/M = f \neq 0$,

$$s^* = k f (1 - f)^{k-1}$$

in which k satisfies the inequalities

$$1/k \leqslant f \leqslant 1/(k - 1)$$

The proof of the converse consists of the following theorems

Theorem 3.3.1

The sum-capacity

$$s^* \triangleq \max_{(s_1,\ldots,s_M)\in S} s = \max_{\tilde{S}} \max_{(s_1,\ldots,s_M)\in \tilde{S}} s$$

The last equality suggests that the sum-capacity may be achieved by some random coding scheme without time sharing. (Each $\tilde{S}$ corresponds to a pure strategy.)

Proof

The sum-throughput s is a linear function of the s_i's. Furthermore, S is the convex hull of the $\tilde{S}$'s. Therefore the maximum of s over S is achieved at the corner point of some $\tilde{S}$.

Q.E.D.

Theorem 3.3.2

$$s^* < \max_{P_1(X_1)\cdots P_M(X_M)} 1/M C_N \sum_{\psi} I(\{X_i\}_{i\in\psi}; Y)$$

Proof

For any $P_i(X_i)$, $1\leq i\leq M$ and any $s_1,\ldots s_M$ in the corresponding $\tilde{S}$, sum those inequalities with $\Omega = \psi$ in theorem 3.1.2, we have

$$\sum_{\psi}\sum_{i\in\psi} s_i < N/M \sum_{\psi} I(\{X_i\}_{i\in\psi}; Y)$$

Using simple counting arguments, the left hand side equals

$$N/M \; {}_M C_N \sum_{i=1}^{M} s_i$$

The theorem follows by substituting this inequality in theorem 3.3.1.

Q.E.D.

What remains for proving the converse is that for large M, the right hand side of the inequality of theorem 3.3.2 becomes

$$\max_{P(X)} I(X^N; Y)$$

Proof

Treating $1/_M C_N$ as the value of the probability distribution $P(\psi)$, we have

$$1/_M C_N \sum_{\psi} I(\{X_i\}_{i\in\psi}; Y)$$

$$= \sum_{\psi} P(\psi)\, I(X_1, \ldots X_M; Y / \psi)$$

$$= I(X_1, \ldots X_M; Y / \Phi)$$

in which Φ is the set of all $\psi \subseteq \{1,\ldots,M\}$ with $/\psi/ = N$. Furthermore,

$$I(X_1, \ldots X_M ; Y/ \Phi)$$

$$= H(Y/ \Phi) - H(Y/\Phi X_1, \ldots X_M)$$

$$\leq H(Y) - H(Y/\Phi X_1, \ldots X_M)$$

Now

$$H(Y/\Phi X_1, \ldots X_M)$$

$$= \sum_{\substack{\psi\in\Phi \\ x_i\in X, \forall i}} P(\psi x_1, \ldots x_M)\, H(Y/\psi x_1, \ldots x_M)$$

The event $\Phi X_1 \ldots X_M$ corresponds to choosing N variables without replacement from the M variables $X_1, \ldots X_M$. For large M, the probability $P(\psi x_1, \ldots x_M)$ can be approximated by the case of choosing with replacement. Thus $H(Y/\Phi X_1 \ldots X_M)$ can be approximated by $H(Y/X^N)$ in which X^N is a set of N independent and identically distributed random variables with probability distribution

$$P(X) = 1/M \sum_{i=1}^{M} P_i(X_i)$$

Therefore

$$I(X_1 \ldots X_M ; Y/\Phi) \leq I(X^N ; Y)$$

Thus we have

$$s^* < \max_{P(X)} I(X^N ; Y)$$

This completes the proof of the converse.

For cases when N is not substantially smaller than M (that is, $N/M = f > 0$), the theorems 3.3.1 and 3.3.2 remain valid. Define $f=N/M$. The remainder of this section derives an upper bound on the sum-throughput for the packet-synchronized collision channel in terms of f. Assuming $P_i(x_i) = p_i$ if x_i is an idle and $P_i(x_i) = (1-p_i)/2^n$ if x_i is a packet, it can be shown that

$$I(\{X_i\}_{i\in\psi} ; Y)$$

$$= \sum_{i\in\psi} I(X_i ; Y/\psi)$$

$$= \sum_{i\in\psi} [(1-p_i)/p_i] \prod_{j\in\psi} p_j \quad \text{n-bit/code symbol}$$

Substituting this value into the inequality of theorem 3.3.2 gives

$$s^* < 1/_M C_N \sum_{\psi} \sum_{i\in\psi} [(1-p_i)/p_i] \prod_{j\in\psi} p_j$$

The right hand side is a multilinear function of the p_i's. Therefore, it has a maximum which occurs when a certain set of the p_i's are 1 while the rest of the p_i's are 0. Since this function is symmetrical in the p_i's, we assume that $p_i=1$ for $1 \leq i \leq k$ and $p_i=0$ for $k+1 \leq i \leq M$. Using simple combinatorial arguments for the right hand side of the above inequality gives

$$s^* < k \; {}_{M-k}C_{N-1} \; / \; {}_{M}C_{N}$$

The maximum occurs for the smallest k satisfying

$$k \; {}_{M-k}C_{N-1} > (k+1) \; {}_{M-(k+1)}C_{N-1}$$

For large M, and f that is not too close to zero, it can be shown from the above inequality that the optimal value of k depends only on the value of f. This fact will be of use later on. Now

$$s < k \; {}_{M-k}C_{N-1} \; / \; {}_{M}C_{N}$$

$$= k \; \{(M-k)!/[(N-1)! \; (M-k-N+1)!]\} \; . \; \{N! \; (M-N)!/ \; M!\}$$

$$= k \; N \; \{(M-k)!/M!\} \; . \; \{(M-N)!/(M-N-k+1)!\}$$

$$= k \; N \; \{1/M^{k}\} \; . \; \{(M-N)^{k-1}\}$$

for large M, when k depends only on f

$$= k \; N/M \; (1 - (N/M)^{k-1})$$

$$= k \; f \; (1-f)^{k-1}$$

For given f, the value of k that maximizes $kf(1-f)^{k-1}$ satisfies

$1/k \leq f \leq 1/(k-1)$

For small f, we have kf=1, hence $s^{*} < e^{-1}$ which agrees with our previous derivation in section 3.1.

The value $kf\,(1-f)^{k-1}$ has a special interpretation, which corresponds to the throughput of the following scheme. The M users are divided into M/k groups, with k users in each group. Each group uses a slot in a round-robin TDMA scheme with M/k slots per frame. The k users in a group would contend for the use of the slot. The throughput of the channel can be shown to be $kf.(1-f)^{k-1}$. This scheme, in fact, bears certain resemblance to the urn scheme proposed by Kleinrock and Yemini [11].

Appendix 3.1 Preambles for user identification

This appendix shows, using a random coding argument, that the amount of preamble required for reliable identification of the set of active users is proportional to log M.

The identification mechanism works as follows. A random codebook, with codewords a_i^r, $1 \le i \le M$, is generated according to the probability distribution $P_a(X)$. The user i, $1 \le i \le M$, is assigned the preamble a_i^r which is transmitted at the beginning of a period if the user has a message to sent. Without loss of generality, we assume that the set of active user is $\mathcal{V} = \{1,2,\ldots,N\}$. Let Y^r be the channel output sequence in the preamble field. The receiver i decides that the transmitter i is active in the period if $(a_i^r, y^r) \in T_\epsilon^r$, in which T_ϵ^r is the set of ϵ-typical r-sequences defined by

$$
\begin{aligned}
T = \{ (x^r,y^r) \in X^r x\, Y^r \\
: 1/r \;\; |-\log P(x^r,y^r) - H(X,Y)| < \epsilon \\
1/r \;\; |-\log P(x^r) - H(X)| < \epsilon \\
1/r \;\; |-\log P(y^r) - H(Y)| < \epsilon \;\; \}
\end{aligned}
$$

For the users in $\mathcal{V}$, failure of identification for user i ($1 \le i \le N$) occurs when $(a_i^r, y^r) \notin T_\epsilon^r$. Over the code ensemble, the probability of failure of identification for user i is bounded by ϵ, for all $\epsilon > 0$ provided that r is larger than some $r_o(N,\epsilon)$ as a result of the law of large numbers. It is noteworthy that r_o

does not depend on M.

For the users not in Ψ, false identification for user i ($N+1 \leq i \leq M$) occurs when $(a_i^r, y^r) \in T_\epsilon^r$. Over the code ensemble, the probability of false identification for user i is

$$
\begin{aligned}
& P\left((A_i^r, Y^r) \in T_\epsilon^r \right) \\
= & \sum_{T_\epsilon^r} P(a_i^r, y^r) && \text{a} \\
= & \sum_{T_\epsilon^r} \prod_{k=1}^{r} P(a_{ik}^r, y^r) && \text{b} \\
= & \sum_{T_\epsilon^r} \prod_{k=1}^{r} P(a_{ik}^r)\, P(y^r) && \text{c} \\
\leq & \sum_{T_\epsilon^r} \exp_2 \left(-r(H(X)-\epsilon) - r(H(Y)-\epsilon)\right) && \text{d} \\
\leq & \exp_2 (r(H(XY)+\epsilon)) \exp_2 \left(-r(H(X)+H(Y)-2\epsilon)\right) && \text{e} \\
= & \exp_2 (-r(I(X;Y)-3\epsilon)) && \text{f}
\end{aligned}
$$

in which b follows from the facts that the preambles are generated independently symbol by symbol and that the channel is memoryless; c is due to the facts that $i \notin \Psi$ and that the preambles are generated independently; the upper bound in d results from the definition of T_ϵ^r; e is due to an upper bound on the cardinality of an ϵ-typical set derived similar to that in chapter 2.

Let $\hat{\Psi}$ be the estimate of Ψ. Over the code ensemble, and applying the union bound, we have

$$P(\hat{\psi} \neq \psi)$$

$$\leq N\epsilon + (M-N)\exp_2(-r(I(X;Y) - 3\epsilon))$$

$$= N\epsilon + \exp_2(-r(I(X;Y) - 1/r \log(M-N) - 3\epsilon))$$

Consequently, the amount of preamble required for reliable identification of the set of active users is proportional to log M.

Chapter 4 The Multiple Access Channel with Incomplete Codebook Knowledge

So far, the multiple access channel considered assumes that the codebooks of all users are known at the receivers. Thus each receiver may determine jointly the most likely code sequences sent by the transmitters. This joint decoding effort, however, is often computationally infeasible. Consequently, the signals of the other users are often treated as interfering noise and are not estimated for noise reduction. For example, joint decoding is not performed for direct sequence spread spectrum multiple accessing. Also for the sake of secrecy, some receivers may not have complete knowledge of the codebooks of the other users. Furthermore, jamming may occur in the system. These circumstances result in incomplete codebook knowledge.

This chapter progressively defines and develops insight into the issue of incomplete codebook knowledge through a series of coding theorems. Section 4.1 studies the reliability of communication for a single user over a memoryless channel with stationary statistics. The decoder may assume a different channel statistics in performing maximum likelihood decoding. The section ends with a theorem which states that reliable communication can be achieved up to I', a new information theoretic quantity which is a function of the codebook statistics, the actual channel statistics and the assumed channel statistics. Section 4.2 gives a proof of this theorem. Section 4.3 investigates the properties (convexity and bounds) of I'.

Section 4.4 considers the model of section 4.2, but with a jammer on the channel. The section starts with a discussion on the type of constraints that may be imposed on the signals sent by the user and the jammer. Using the result of section 4.2, we obtain a coding theorem about what the user and the jammer can do respectively to ensure or impede reliable communication. Section 4.5 examines the result of section 4.4 in an asynchronous multiple accessing environment.

The results in this chapter are new in the following aspects. Unlike mutual information in classical information theory, the information theoretic quantities in this chapter take into account the structure of the decoder. Joint decoding is restricted by imposing a specific decoder structure. The decoder with incorrect knowledge of the channel statistics was introduced in [13], but without a detailed study. The result of section 4.2 is new and provides significant insights into the problem. The result of the single user single jammer channel of section 4.4 appears to be similar to that of Blackwell , Breiman and Thomasian [12], Stiglitz [13] as well as McEliece and Stark [14]. In [12], the unknown channel is assumed memoryless and has stationary statistics with the channel transition probability distribution in a certain permissible set P_c. In [13], the unknown channel is assumed memoryless but may have non-stationary statistics. The channel transition probability distribution, which may vary from instance to instance, is in the permissible set P_c. Both [12] and [13] assume that the input probability distribution for the code symbols for the user is in a

permissible set P_s. Both use a mini-max argument and arrive at the conclusion that the channel capacity for the unknown channel is

$$C = \min_{P(Y/X)\in P_c} \max_{P(X)\in P_s} I(X;Y)$$

There are two objections to their models in the presence of jamming. First, unlike the unknown channel modeled in [12] and [13], a channel with a jammer may not be memoryless or have stationary statistics. Second, coding or jamming strategies are usually constrained on a per codeword basis (that is, the total cost for the symbols of each code sequence or jamming sequence must be less than a certain amount) rather than on a per symbol basis. While section 4.3 arrives at a similar statement on the capacity of the channel, the formulation does not require memorylessness of the channel and adopts constraints for signaling on a per code sequence basis. The proof of the coding theorem gives new insights on how to choose the best decoding metric to for various jamming strategies. The work of McEliece and Stark [14] does not require memorylessness and stationary statistics for the user and the jammer. The problem they considered involves a two party game for mutual information between the transmitter and the jammer and it is not clear how a coding theorem may follow from such a framework.

4.1 Communication with incorrect model of a memoryless channel

This section gives an achievable throughput of a channel with a decoder which has an incorrect model of the channel. Before stating the result, we would like to examine the meaning of " an incorrect model of the channel for the decoder".

The communication system consists of a single source which sends one of the the 2^{nR} equiprobable mesages j, $1 \leq j \leq 2^{nR}$. Each message j is encoded as $x^n(j)$, a sequence of n code symbols from the alphabet $X=\{1,2,\ldots,U\}$. The channel puts out y^n, a sequence of n channel symbols from the alphabet $Y=\{1,2,\ldots,W\}$, according to the channel transition probabilities $P(y^n/x^n)$. The decoder assumes that the channel transition probabilities are $P'(y^n/x^n)$. The decoder performs maximum likelihood (ML) decoding according to the assumed channel transition probabilities. In other words, the decoder chooses the estimate $\hat{j}$ such that

$$P'(y^n/x^n(\hat{j})) > P'(y^n/x^n(j))$$

for all j. We shall restrict ourselves to memoryless channels in this section. More generally, the ML decoder may assume a metric a_{xy} between each code symbol $x \in X$ and each channel symbol $y \in Y$. Let n_{xy} be the number of k's such that $x_k^n(j)=x$ and $y_k^n=y$. (The subscript k denotes the k-th letter of a sequence.) The ML decoder with $A=\{a_{xy}\}$ gives the estimate $\hat{j}$ which maximizes the sum-metric

$$\sum_{xy} a_{xy} n_{xy}$$

for all j.

Denoting P(X=x), P(Y=y) and P(X=x,Y=y) by p_x, p_y and p_{xy} respectively, we have the following theorem

Theorem 4.1

Reliable communication is achievable if the rate of encoding R is less than

$$C = \max_{P(X)} I'(X;Y) \quad \text{(nats)}$$

in which

$$I'(X;Y) = H(X) + H(Y) - H'(XY)$$

$$H(X) = \sum_x -p_x \ln p_x$$

$$H(Y) = \sum_y -p_y \ln p_y$$

and

$$H'(XY) = \max_F H(F) = \max_F \sum -f_{xy} \ln f_{xy} \qquad (F \triangleq \{f_{xy}\})$$

subject to

$f_{xy} \geq 0$ for all x,y

$\sum_y f_{xy} = \sum_y p_{xy}$ for all x

$\sum_x f_{xy} = \sum_x p_{xy}$ for all y

$\sum_{xy} f_{xy} a_{xy} \geq \sum_{xy} p_{xy} a_{xy}$

This theorem is proved in the next section.

The information theoretic term I' is analogous to mutual information in classical information theory defined by

$$I(X;Y) = \sum_{xy} p_{xy} \ln (p_{xy} / [p_x p_y])$$

The difference is that I' is evaluated at the f_{xy}'s (instead of the p_{xy}'s) which minimize the value of I, subject to a set of linear contraints involving p_{xy} and a_{xy}. C is the maximum value of I' for various input distributions P(X). The properties of I' and H' will be studied in section 4.3.

There is reason to believe that C is indeed the capacity of the channel with a decoder that assumes the metric A. Further work is required to prove this conjecture.

4.2 Proof for achievability

We show that random coding can achieve reliable communication for information rate R less than C. Let the 2^{nR} codewords of the codebook C be chosen independently codeword by codeword and letter by letter according to the probability distribution P(X). Furthermore, we require each codeword $x^n(j)$ to be in the set

$$T^n_{\epsilon,x} = \{x^n : n(p_x - \epsilon) \leq n_x \leq n(p_x + \epsilon) \text{ for all } x\}$$

in which n_x is the number of letters in x^n that equals x. Essentially, each codeword must have a typical number of each letter in the input alphabet. Let $\mathbf{C}$ be the ensemble of such C's.

For the sake of symmetry and obtaining a tight bound, the channel sequence y^n is observed by the decoder only if y^n is in the set

$$T^n_{\epsilon,y} = \{y^n : n(p_y - \epsilon) \leq n_y \leq n(p_y + \epsilon) \text{ for all } y\}$$

in which n_y is the number of letters in y^n that equal y, and $p_y = p_x . P(y/x)$. The decoder declares a decoding failure when y^n is not in $T^n_{\epsilon,y}$. It should be noted that the decoder cannot decide whether y^n is in $T^n_{\epsilon,y}$, since it does not know P(y/x). We assume that a genie would "black-out" the channel if y^n is not in $T^n_{\epsilon,y}$. Obviously, the genie makes the probability of decoding error worse.

The ensemble error probability is

$$P_{\mathcal{C}}(\hat{J}\neq J) = \sum_j P(j)\, P_{\mathcal{C}}(\hat{J}\neq j) = P_{\mathcal{C}}(\hat{J}\neq j)$$

with the last equality due to the symmetry of the random code in the generation of each codeword.

Consider the threshold decoder which finds all $\tilde{j}$ such that $(x^n(\tilde{j}), y^n) \in T^n_{\epsilon,+}$. where

$$T^n_{\epsilon,+} = \{ (x^n,y^n): \sum_{xy} n_{xy}\, a_{xy} \geq n \left(\sum_{xy} p_{xy}\, a_{xy} - \epsilon \right) \}$$

(Note that the threshold decoder cannot be implemented since we do not know P(y/x) which determines the value of the threshold.) If no such $\tilde{j}$ or more than one such $\tilde{j}$ exist, the threshold decoder declares a decoding failure. Otherwise, the decoder would give the estimate $\hat{j}=\tilde{j}$. Obviously, the output of the threshold decoder equals that of the ML decoder (which is implemented and chooses the $\hat{j}$ with the maximum sum-metric) when the threshold decoder does not declare a decoding failure. Consequently, the error probability of the ML decoder is upper bounded by that of the threshold decoder. Therefore, the ensemble error probability $P_{\mathcal{C}}(\hat{J}\neq j)$ is upper bounded by

$$P_{\mathcal{C}}\Big(\{(X^n(j),Y^n) \notin T^n_{\epsilon,+}\} \cup \{Y^n \notin T^n_{\epsilon,y}\} \cup \{\text{some } \tilde{j}\neq j: (X^n(\tilde{j}),Y^n) \in T^n_{\epsilon,+},\ Y^n \in T^n_{\epsilon,y}\} \Big)$$

$$\leq P_{\mathcal{C}}\left((X^n(j),Y^n) \notin T^n_{\epsilon,+} \right) + P_{\mathcal{C}}\left(Y^n \notin T^n_{\epsilon,y}\right) + \sum_{\tilde{j}\neq j} P_{\mathcal{C}}\left((X^n(\tilde{j}),Y^n) \in T^n_{\epsilon,+},\ Y^n \in T^n_{\epsilon,y} \right)$$

The inclusion of $Y^n \in T^n_{\epsilon,y}$ in the last term tightens the upper bound, as we shall see later. The first two terms are each less than ϵ provided n is larger than some n_0, which is shown in appendix 4.1 using the law of large numbers. The last term equals

$$\sum_{\tilde{j}\neq j} \sum_{\substack{(x^n(\tilde{j}),y^n): \\ (x^n(\tilde{j}),y^n)\in T^n_{\epsilon,xy} \\ y^n \in T^n_{\epsilon,y}}} P(x^n(\tilde{j})\, y^n)$$

$$= \sum_{\tilde{j}\neq j} \sum_{\substack{(x^n(\tilde{j}),y^n): \\ (x^n(\tilde{j}),y^n)\in T^n_{\epsilon,xy} \\ y^n \in T^n_{\epsilon,y}}} P_{\mathcal{C}}(x^n(\tilde{j})).P_{\mathcal{C}}(y^n) \qquad \text{a}$$

$$= \sum_{\tilde{j}\neq j} \sum_{K} P_{\mathcal{C}}(x^n(\tilde{j})).P_{\mathcal{C}}(y^n) \qquad \text{b}$$

where $K=\{(x^n,y^n) : (x^n,y^n)\in T^n_{\epsilon,xy},\ x^n\in T^n_{\epsilon,x},\ y^n\in T^n_{\epsilon,y}\}$

$$\leq \sum_{\tilde{j}\neq j} \sum_{K} \exp\left(-n(H(X)-\epsilon)\right) \exp\left(-n(H(Y)-\epsilon)\right) \qquad \text{c}$$

$$\leq |K| \exp_2(nR) \exp\left(-n(H(X)-\epsilon)\right) \exp\left(-n(H(Y)-\epsilon)\right) \qquad \text{d}$$

(Note: Throughout this chapter, exp is the natural exponential and the entropy H is defined using natural logarithm.) The product in a is due to the fact that y^n is independent of $x^n(\tilde{j})$ if $\tilde{j}\neq j$ since different codewords are chosen independently; b results from the fact that the construction of the random code requires $x^n(\tilde{j})$ to be in $T^n_{\epsilon,x}$, hence $P_{\mathcal{C}}(x^n(\tilde{j}))=0$ if $x^n(\tilde{j})\notin T^n_{\epsilon,x}$; c follows from an upper bound derived in appendix 4.2 on $P_{\mathcal{C}}(x^n)$ and

$P_{\mathcal{C}}(y^n)$ for all $x^n \in T^n_{\epsilon,x}$ and $y^n \in T^n_{\epsilon,y}$; d follows from the fact that there are 2^{nR} codewords.

The key of the proof of this section is the following upper bound on the cardinality of the set K.

Theorem 4.2

For all $\epsilon > 0$, there exists some n_0 such that if $n > n_0$, then

$$|K| \leq \exp\left(n(H'(XY) + \epsilon)\right)$$

in which

$$H'(XY) = \max_{F} \quad H(F) \qquad \left(F \triangleq \{f_{xy}\} \right)$$

subject to

$$f_{xy} \geq 0 \qquad \text{for all } x,y$$

$$\sum_{y} f_{xy} = \sum_{y} p_{xy} \qquad \text{for all } x$$

$$\sum_{x} f_{xy} = \sum_{x} p_{xy} \qquad \text{for all } y$$

$$\sum_{xy} f_{xy} a_{xy} \geq \sum_{xy} p_{xy} a_{xy}$$

Proof

By the definition of K, $(x^n, y^n) \in K$ if

$$x^n \in T^n_{\epsilon,x} \iff n(p_x - \epsilon) \leq n_x \leq n(p_x + \epsilon) \quad \text{for all } x \qquad 4.2.1a$$

$$y^n \in T^n_{\epsilon,y} \iff n(p_y - \epsilon) \leq n_y \leq n(p_y + \epsilon) \quad \text{for all } y \qquad b$$

$$(x^n, y^n) \in T^n_{\epsilon,t} \iff n\left(\sum_{xy} p_{xy}\, a_{xy} - \epsilon\right) \leq \sum_{xy} n_{xy}\, a_{xy} \qquad c$$

$$\sum_{xy} n_{xy} = n \qquad d$$

Let $\{n_{xy}\}$ be a set of n_{xy} for all $x \in X$ and $y \in Y$ such that 4.2.1 is satisfied. Let N be the set of all possible $\{n_{xy}\}$. For each $\{n_{xy}\}$, there are

$$n! \,/\, \prod_{xy} (n_{xy})!$$

different (x^n, y^n). Consequently

$$|K| = \sum_{N} n! \,/\, \prod_{xy} (n_{xy})!$$

$$\leq N \max_{N} n! \,/\, \prod_{xy} (n_{xy})!$$

The number of partitions of n into the $\{n_{xy}\}$'s satisfying 4.2.1d equals

$$n+UW-1\, C_{UW-1} \leq n^{UW}$$

Thus

$$|N| \leq n^{UW}$$

On the other hand

$$n! \,/\, \prod_{xy} (n_{xy})!$$

$$= \binom{n}{n_{11}} \binom{n-n_{11}}{n_{12}} \binom{n-n_{11}-n_{12}}{n_{13}} \cdots \binom{n-n_{11}-n_{12}-\cdots-n_{ij}}{n_{i(j+1)}} \cdots \binom{n-n_{11}-n_{12}-\cdots n_{U(W-1)}}{n_{UW}} \quad a$$

$$\leq \exp(n\,H(n_{11}/n)).\exp((n-n_{11})\,H(n_{12}/(n-n_{11}))).$$

$$\exp((n-n_{11}-n_{12})\,H(n_{13}/(n-n_{11}-n_{12}\quad)))\ldots..$$

$$\exp((n-n_{11}-n_{12}-\ldots-n_{ij})\,H(n_{i(j+1)}/(n-n_{11}-n_{12}\ldots-n_{ij})))\ldots.$$

$$\exp((n-n_{11}-n_{12}-\ldots-n_{U(W-1)})\,H(n_{UW}/(n-n_{11}-n_{12}\ldots-n_{U(W-1)}))) \qquad b$$

where $H(x) = -x \ln x - (1-x) \ln (1-x)$

$$= \exp(n\,\{H(f_{11})+(1-f_{11})H(f_{12}/(1-f_{11}))+$$

$$(1-f_{11}-f_{12})H(f_{13}/(1-f_{11}-f_{12}))+\ldots..+$$

$$(1-f_{11}-f_{12}-\ldots-f_{ij})H(f_{i(j+1)}/(1-f_{11}-f_{12}-\ldots-f_{ij}))+\ldots+$$

$$(1-f_{11}-f_{12}-\ldots-f_{U(W-1)})H(f_{UW}/(1-f_{11}-f_{12}-\ldots-f_{U(W-1)}))\}) \qquad c$$

$$= \exp(n\,H(F)) \qquad d$$

where $H(F) = \sum_{xy} -f_{xy} \ln f_{xy}$

in which a follows from a rearrangement of the product; b follows from a bound on binormial coefficients (part b of problem 5.8 in [9]); c introduces the f_{xy}'s; and d follows from simple algebraic manipulation.

Dividing all the inequalities in 4.2.1 by n, and choosing a small enough ϵ give the constraints

$$f_{xy} \geq 0 \qquad \text{for all } x,y \qquad 4.2.2a$$

$$\sum_y f_{xy} = \sum_y p_{xy} \qquad \text{for all } x \qquad \text{b}$$

$$\sum_x f_{xy} = \sum_x p_{xy} \qquad \text{for all } y \qquad \text{c}$$

$$\sum_{xy} f_{xy}\, a_{xy} \geq \sum_{xy} p_{xy}\, a_{xy} \qquad \text{d}$$

Consequently, the cardinality of K is upper bounded by

$$n^{UW} \exp\left(n \max_{F \text{ satisfying } 4.2.2} H(F)\right)$$

$$= \exp\left(n \left(\max_{F \text{ satisfying } 4.2.2} H(F) + UW (\ln n) / n \right)\right)$$

Thus for every $\epsilon > 0$, there exist some n_0 satisfying

$$UW (\ln (n_0 - 1)) / (n_0 - 1) < \epsilon < UW (\ln n_0) / n_0$$

such that

$$|K| < \exp (n H'(XY) + \epsilon)$$

for all $n > n_0$.

Q.E.D.

Summarizing the results so far, we have

$$P_e(\hat{J} \neq j) < 2\epsilon + \exp (-n (H(X) + H(Y) - H'(XY) - R - 3\epsilon))$$

$$= 2\epsilon + \exp (-n (I'(X;Y) - R - 3\epsilon))$$

(R in nats.) Therefore, the ensemble error probability can be made arbitrarily small for large n provided $R < I'(X;Y) - 3\epsilon$. This completes the proof for theorem 4.1.

4.3 Properties of I' and H'

This section derives convexity properties and bounds on H'(XY) and I'(X;Y), which are defined by

$$H'(XY) = \max_{F} \sum_{xy} -f_{xy} \ln f_{xy} \qquad 4.3.1$$

$$I'(X:Y) = \min_{F} \sum_{xy} f_{xy} \ln \left(f_{xy} / \left(\sum_{y} f_{xy} \sum_{x} f_{xy} \right)\right) \qquad 4.3.2$$

both subject to the same set of constraints on F as follows

$$f_{xy} \geq 0 \qquad \text{for all } x,y \qquad 4.3.3a$$

$$\sum_{y} f_{xy} = \sum_{y} p_{xy} \qquad \text{for all } x \qquad b$$

$$\sum_{x} f_{xy} = \sum_{x} p_{xy} \qquad \text{for all } y \qquad c$$

$$\sum_{xy} f_{xy} a_{xy} \geq \sum_{xy} p_{xy} a_{xy} \qquad d$$

Theorem 4.3.1

$$H(X) + H(Y) \geq H'(XY) \geq H(XY)$$

$$0 \leq I'(X;Y) \leq I(X;Y)$$

Proof

Ignoring the constraint 4.3.3d, we have

$$H'(XY) \leq \max_{F} \sum_{xy} -f_{xy} \ln f_{xy}$$

subject to

$f_{xy} \geq 0$ for all x,y

$f_x = p_x$ for all x $(f_x \triangleq \sum_y f_{xy})$

$f_y = p_y$ for all y $(f_y \triangleq \sum_x f_{xy})$

Consequently

$$\sum_{xy} -f_{xy} \ln f_{xy}$$

is maximized when $f_{xy}=f_x . f_y=p_x . p_y$ for all x, y. Thus

$$H'(XY) \leq H(X) + H(Y)$$

On the other hand, letting $f_{xy} = p_{xy}$ satisfies all constraints in 4.3.3 and gives

$$\sum_{xy} -f_{xy} \ln f_{xy} = \sum_{xy} -p_{xy} \ln p_{xy} = H(XY)$$

Consequently

$$H'(XY) = \max_F \sum_{xy} -f_{xy} \ln f_{xy} \geq H(XY)$$

The inequalities $0 \leq I'(X;Y) \leq I(X;Y)$ follows from the definition

$$I'(X;Y) = H(X) + H(Y) - H'(XY)$$

Q.E.D.

The non-negativity of I' is very pleasing. In contrast, the quantity

$$I^* \triangleq \sum_{xy} p_{xy} \ln (p'_{xy} / (p'_x p'_y))$$

(where p_{xy} is the actual probability and p'_{xy} is the assumed probability) satisfies $I^* \leq I$ but does not possess the non-negative property.

Theorem 4.3.2

$$\min_{A} \quad H'(XY) = H(XY)$$

$$\max_{A} \quad I'(X;Y) = I(X;Y)$$

which are achieved when $a_{xy} = \ln p_{xy}$.

Proof

From the previous theorem, we know that $H'(XY) \geq H(XY)$ for all A. We shall show that for $A=\{\ln p_{xy}\}$, $H'(XY)=H(XY)$. In the maximization of $\sum_{xy} -f_{xy} \ln f_{xy}$, we have

$$\sum_{xy} f_{xy} \ln 1/f_{xy}$$

$$\leq \sum_{xy} f_{xy} \ln 1/p_{xy}$$

$$\leq \sum_{xy} p_{xy} \ln 1/p_{xy}$$

$$= H(XY)$$

in which the first inequality follows from the well-known identity [9] that

$$\sum_i \; p_i \ln (q_i/p_i) \leq \sum_i \; p_i (q_i/p_i - 1) = 0$$

and the second inequality follows from the constraint 4.3.3d

$$\sum_{xy} \; f_{xy} \, a_{xy} \geq \sum_{xy} p_{xy} \, a_{xy}$$

Since letting $f_{xy}=p_{xy}$ satisfies the constraints in the definition of H'(XY) and gives

$$\sum_{xy} \; -f_{xy} \ln f_{xy} = H(XY)$$

we have

$$\min_A \; H'(XY) = H(XY)$$

Q.E.D.

Theorem 4.3.3

H'(XY) is convex $\cap$ in the probability distribution $\{p_{xy}\}$.

Proof

Let $F_.^a$ be an $\{f_{xy}^a\}$ satisfying the constraints in (4.3.3) for $P_.^a=\{p_{xy}^a\}$. (The index a is a label). Let $\lambda F_.^1+(1-\lambda)F_.^2=\{\lambda f_{xy}^1+(1-\lambda) f_{xy}^2\}$. Let $\bar{F}_.^1$ and $\bar{F}_.^2$ be the $F_.^1$ and $F_.^2$ that achieves the maximum of $H(F_.^1)$ and $H(F_.^2)$ respectively. Thus

$$\lambda \max_{F^1} H(F^1) + (1-\lambda) \max_{F^2} H(F^2)$$

$$\leq H(\lambda \bar{F}^1 + (1-\lambda)\bar{F}^2)$$

since H(F) is a convex $\cap$ function of F [9]. Now $\lambda\bar{F}^1+(1-\lambda)\bar{F}^2$ satisfies the constraints in (4.3.3) for $P^{\lambda} \triangleq \{\lambda p'_{xy}+(1-\lambda)p^2_{xy}\}$, consequently

$$H(\lambda \bar{F}^1 + (1-\lambda)\bar{F}^2) \leq \max_{F^{\lambda}} H(F^{\lambda})$$

Q.E.D.

As a corollary, H'(XY) is a convex $\cap$ function of $\{p_x\}$ for fixed $\{p_{y|x}\}$. This follows from the fact that $p_{xy}=p_x \cdot p_{y|x}$ is a linear function of p_x for given $p_{y|x}$. Similarly, H'(XY) is a convex $\cap$ function of $\{p_{y|x}\}$ for fixed $\{p_x\}$.

Theorem 4.3.4

I'(X;Y) is a convex U function of $\{p_{y|x}\}$ for fixed $\{p_x\}$.

Proof

The reasoning is similar to that of theorem 4.3.3. Define $f_{y|x}=f_{xy}/f_x$. Let $F^{a}_{y|x}$ be an $\{f^{a}_{y|x}\}$ satisfying the constraints in (4.3.3) for $P_{y|x}=\{p^{a}_{y|x}\}$, with fixed $f_x=p_x$ for all x. Let $\bar{F}^1_{y|x}$ and $\bar{F}^2_{y|x}$ denote the $F^1_{y|x}$ and $F^2_{y|x}$ that achieves the minimum of $I(F^1_{y|x})$ and $I(F^2_{y|x})$, with

$$I(F_{y|x}) \triangleq \sum_{xy} f_x f_{y|x} \ln \left(f_{y|x} / \left(\sum_x f_{y|x} f_x\right)\right)$$

It is well known [9] that $I(F_{y|x})$ is a convex U function of $\{f_{y|x}\}$ for fixed $\{f_x\}$. Consequently

$$\lambda \min_{F^1_{y|x}} I(F^1_{y|x}) + (1-\lambda) \min_{F^2_{y|x}} I(F^2_{y|x})$$

$$\geq I(\lambda F^1_{y|x} + (1-\lambda) F^2_{y|x})$$

$$\geq \min_{F^\lambda_{y|x}} I(F^\lambda_{y|x})$$

in which $F^\lambda_{y|x}$ satisfies the constraints in (4.3.3) for $P^\lambda_{y|x}=\{\lambda p^1_{y|x}+(1-\lambda)p^2_{y|x}\}$

Q.E.D.

For the binary input binary output channel (that is $X=Y=\{0,1\}$), H' and I' can be easily found. Let the actual probabilities be p_{xy} and the assumed probabilities be p'_{xy} (that is $a_{xy}=\ln p_{xy}$). We have the following result.

Theorem 4.3.5

For the binary input binary output channel, $H'(XY)=H(XY)$ (consequently $I'(X;Y)=I(X;Y)$) if $(p_{11}p_{00})/(p_{10}p_{01})$ and $(p'_{11}p'_{00})/(p'_{10}p'_{01})$ are both larger than one or both smaller than 1. Otherwise, $H'(XY)=H(X)+H(Y)$ (consequently $I'(X;Y)=0$).

Proof

The maximization for H'(XY) involves four equality constraints (from 4.3.3b and c), one of which is redundant, and the

inequality constraint of 4.3.3d. There are four unknowns, namely the f_{xy}'s. Without the inequality constraint 4.3.3d, H'(XY) is maximized with

$$f_{xy} = \left(\sum_{y'=0,1} p_{xy'} \right) \left(\sum_{x'=0,1} p_{x'y} \right) \qquad *$$

giving H'(XY)=H(X)+H(Y). We shall check whether these particular values of $\{f_{xy}\}$ verify the inequality constraint. Substituting the values of f_{xy} in * into the constraint, we have

$$\sum_{xy} (f_{xy} - p_{xy}) a_{xy}$$

$$= \sum_{xy} \{ \left(\sum_{y'=0,1} p_{xy'} \right) \left(\sum_{x'=0,1} p_{x'y} \right) - p_{xy} \} \ln p'_{xy}$$

Consider the term for x=0 and y=0 in the outermost summation.

$$\{ (p_{00} + p_{01}) (p_{00} + p_{10}) - p_{00} \} \ln p'_{00}$$

$$= \{ p_{00} (p_{00} + p_{01} + p_{10}) + p_{01}p_{10} - p_{00} \} \ln p'_{00}$$

$$= \{ p_{00} (1 - p_{11}) + p_{01}p_{10} - p_{00} \} \ln p'_{00}$$

$$= \{ p_{01}p_{10} - p_{00}p_{11} \} \ln p'_{00}$$

Similarly, the other terms become $-\{p_{01}p_{10}-p_{00}p_{11}\}\ln p'_{01}$, $-\{p_{01}p_{10}-p_{00}p_{11}\}\ln p'_{10}$, and $\{p_{01}p_{10}-p_{00}p_{11}\}\ln p'_{11}$. Consequently,

$$\sum_{xy} (f_{xy} - p_{xy}) a_{xy}$$

$$= \{p_{01}p_{10} - p_{00}p_{11}\} \ln [(p'_{00}p'_{11})/(p'_{01}p'_{10})]$$

which is larger than zero if $(p'_{00}p'_{11})/(p'_{01}p'_{10})>1$ and $(p_{00}p_{11})/(p_{01}p_{10})<1$, or $(p'_{00}p'_{11})/(p'_{01}p'_{10})<1$ and $(p_{00}p_{11})/(p_{01}p_{10})>1$. Otherwise, the maximum of $\sum_{xy} -f_{xy} \ln f_{xy}$

occurs when the inequality constraint is satisfied with equality, because the set of f_{xy} satisfying the constraints is convex and the function $\sum_{xy} -f_{xy} \ln f_{xy}$ is a convex $\cap$ function of the f_{xy}'s. For such case, letting $f_{xy}=p_{xy}$ trivially satisfies the constraints 4.3.3a-d (with the constraint 4.3.3d being an equality). This solution is unique since there are four unknowns in four linearly independent equations. Consequently H'(XY)=H(XY).

Q.E.D.

For an U inputs W outputs channel, there are UW unknowns constrained by U+W equality constraints (U+W-1 of which are linearly independent) and one inequality constraint (not counting the non-negative constraints on the f_{xy}'s). If the inequality constraint is verified for

$$f_{xy} = \left(\sum_{y'} p_{xy'} \right) \left(\sum_{x'} p_{x'y} \right)$$

then H'(XY)=H(X)+H(Y) and I'(X;Y)=0. Otherwise, $\sum_{xy} -f_{xy} \ln f_{xy}$ is maximized over an UW-(U+W)-1 dimensional space. Thus H'(XY) can have any value in between H(X)+H(Y) and H(XY).

4.4 Communication in the presence of jamming

There are three types of constraints that may be imposed on the user and the jammer. A per codeword constraint imposes an upper bound on the cost of each codeword, that is

$$b(x^n) = \sum_k b(x_k^n) \leq n B$$

A average codeword constraint imposes an upper bound on the average cost of transmission, that is

$$\sum_j P(j)\, b(x^n(j)) \leq n B$$

A per letter input constraint on a probability distribution $P(X)$ is given by

$$\sum_x P(x)\, b(x) \leq B$$

The per codeword constraint is stronger than the average codeword constraint and is more sensible practically. We shall show how these three constraints are related in proving the following coding theorem. We assume the channel is memoryless with stationary statistics. For the sequences x^n and z^n sent by the user and the jammer respectively, and the channel sequence y^n

$$P(y^n/x^n z^n) = \prod_k P(y_k^n /x_k^n z_k^n)$$

Furthermore, the decoder assumes a certain set of $\{a_{xy}\}$ for decoding. The direct part and the converse are stated respectively as

Theorem 4.4.1 Direct part

Assume that each codeword $x^n(j)$ and the jamming sequence z^n satisfy the per codeword constraints

$$b_X(x^n) \leq n\, B_X$$

and

$$b_Z(z^n) \leq n\, B_Z$$

Reliable communication for the user is achievable if

$$R < \min_{P(Z)\in\mathcal{Z}} \max_{P(X)\in\mathcal{X}} I(X;Y)$$

in which $\mathcal{X}$ is the set of probability distributions on X satisfying

$$\sum_{x\in X} P(x)\, b_X(x) \leq B_X$$

and Z is the set of probability distributions on Z satisfying

$$\sum_{z\in Z} P(z)\, b_Z(z) \leq B_Z$$

Theorem 4.4.2 The converse

Assume that the codeword sequences $x^n(j)$ and the jamming sequence z^n satisfy the average codeword constraints

$$\sum_j P(j)\, b_X(x^n(j)) \leq n\, B_X$$

$$\sum_{z^n} P(z^n)\, b_Z(z^n) \leq n\, B_Z$$

the jammer can make the bit error probability of the user bounded above zero if

$$R > \min_{P(Z)\in\mathcal{Z}} \max_{P(X)\in\mathcal{X}} I(X;Y)$$

in which $\mathcal{Z}$ and $\mathcal{X}$ are the same as that of theorem 4.4.1

The direct part and the converse differs mainly in the type of constraints placed on the sequences x^n and z^n. We shall use random coding for the jamming sequence, according to the probability distribution $P(Z)\in\mathcal{Z}$. We do not use the per codeword constraint for the converse for the following reasons. In proving the converse, the error probability has to be lower bounded for all n. For small n, it is likely that $b_Z(z^n)$ may exceed nB_Z on a per codeword basis. This problem disappears if we use an average codeword constraint instead. Second, the per codeword constraint poses a diophantine problem for small n as illustrated in the following example. Consider n=1, Z={0,1}, $b_X(0)=0$, $b_X(1)=1$ and $B_X=1/4$. Consequently, the only sequence satisfying the per codeword constraint is 0. Therefore, the converse would not be valid if it assumes a per codeword constraint instead. Third, the sequence z^n which satisfies the per codeword constraint is not memoryless, and the memoryless property is required in proving the converse. Note that for large n, there is little difference between using the per codeword constraint and the average codeword constraint.

Proof of the converse

For the converse, z^n is generated independently letter by letter according to the probability distribution $P(Z) \in \mathcal{Z}$. At the decoder, we assume that a genie informs the decoder about P(Z) so that the decoder would use the best metric. Consequently, the channel with the jammer is memoryless and has known and stationary statistics to the decoder. For given P(Z), it is well known that the error probability of the user is lower bounded above zero if

$$R > \max_{P(X)} I(X;Y)$$

By choosing the P(Z) that minimizes the right hand side of the above inequality, the bit error probability of the user is lower bounded above zero if

$$R > \min_{P(Z)\in\mathcal{Z}} \max_{P(X)\in\mathcal{X}} I(X;Y)$$

This completes the proof of the converse

Q.E.D.

Proof of the direct part

For the direct part, we want to show that some codes with large n achieve an almost zero error probability for all z^n that satisfy the per codeword constraint. The decoder is the ML decoder described in section 4.1, which uses the decoding metric $\{a_{xy}\}$. However, the received sequence y^n for a channel with a jammer may

not be memoryless or has stationary statistics, in contrast with the memoryless channel with stationary statistics considered in section 4.1. For each z^n, define the probability distribution $P(Z)$ such that $P(Z=z)=n_z/n$ (n_z is the number of z occuring in z^n). Since z^n satisfies the per codeword constraint $b_Z(z^n)<nB_Z$, we have $P(Z)\in\mathcal{Z}$. We generate the codewords x^n by random coding according to the probability distribution $P(X)\in\mathcal{X}$. For large n, the probability that a randomly generated codeword would not satisfy the per codeword constraint is negligibly small. The direct part shows that for large n, the ensemble error probability is almost zero for each z^n satisfying the per codeword constraint, provided that the rate of transmission is less than the capacity.

Similar to section 4.1, the decoder uses a set of metric $\{a_{xy}\}$ and performs maximum likelihood decoding according to the assumed metric. Since both the random coding scheme and the decoding scheme are symmetrical for each location of the n-sequences, the ensemble error probability is invariant for different permutations of z^n. Consequently, the ensemble error probability for a given z^n is the same as the ensemble error probability over random permutations of z^n. Under random permutations of z^n, the channel with the jammer may be treated as memoryless[@] and has stationary statistics with channel transition probabilities given by $P(y/x) = \sum_z P(z)\ P(y/xz)$,

---

@ In fact, theorem 4.1 remains valid for channels with memory, with $P(y/x)$ denoting the time-averaged transition probabilities.

with P(Z) as defined in the previous paragraph. The result of section 4.2 is applicable and the ensemble error probability for large n is very small if $R<I'(X;Y)$.

Since I' is a function of P(X), P(Z) and $\{a_{xy}\}$ (abbreviated as P_X, P_Z and A respectively) we shall denote $I'(X;Y)$ by $G(P_X,P_Z,A)$ instead. The remainder of this section investigates $G(P_X,P_Z,A)$ for various strategies of choosing P_X and A by the user and P_Z by the jammer. The user would try to choose P_X and A to maximize the minimum of G for all strategies P_Z chosen by the jammer. The jammer would try to choose P_Z to minimize the maximum of G for all strategies P_X and A chosen by the user. Thus proving the direct part is equivalent to proving the following theorem.

Theorem 4.4.3

$$\max_{P_X \in \mathcal{X},\, A} \; \min_{P_Z \in \mathcal{Z}} \; G(P_X,P_Z,A)$$

$$= \min_{P_Z \in \mathcal{Z}} \; \max_{P_X \in \mathcal{X},\, A} \; G(P_X,P_Z,A)$$

$$= G(P_X^*,P_Z^*,A^*)$$

$$= \min_{P_Z \in \mathcal{Z}} \; \max_{P_X \in \mathcal{X}} \; I(X;Y)$$

in which P_X^* and P_Z^* achieve the mini-max of $I(X;Y)$ and $A^*=\{a_{xy}^*\}=\{\ln p_{xy}^*\}$

Proof

By the saddle point theorem [18],

$$\max_{P_X \in \mathcal{X}, A} \min_{P_Z \in \mathcal{Z}} G(P_X, P_Z, A) = \min_{P_Z \in \mathcal{Z}} \max_{P_X \in \mathcal{X}, A} G(P_X, P_Z, A)$$

if and only if there exists some (P_X^*, P_Z^*, A^*) such that for all P_X, A and P_Z,

$$G(P_X, P_Z^*, A)$$

$$\leq \max_{P_X' \in \mathcal{X}, A'} G(P_X', P_Z^*, A') \qquad 4.3.1a$$

$$= G(P_X^*, P_Z^*, A^*) \qquad b$$

$$= \min_{P_Z' \in \mathcal{Z}} G(P_X^*, P_Z', A^*) \qquad c$$

$$\leq G(P_X^*, P_Z, A^*) \qquad d$$

By theorem 4.3.2, with $A^* = \{\ln p_{X|y}^*\}$, we have

$$\max_{P_X' \in \mathcal{X}} \max_{A'} G(P', P^*, A') = \max_{P_X' \in \mathcal{X}} I(X;Y)\Big|_{P(Z)=P_Z^*}$$

Furthermore

$$\max_{P_X' \in \mathcal{X}} I(X;Y)\Big|_{P(Z)=P_Z^*} \geq I(X;Y)\Big|_{P(Z)=P_Z^*} \geq G(P_X, P_Z^*, A)$$

for all P_X, A, with the last inequality due to theorem 4.3.2. This establishes 4.3.1a and b.

4.3.1c and d can be interpreted as follows. If the decoding metric is optimal for the optimal jamming strategy, then the decoder using the same metric would perform better (in the sense that $G(P_X, P_Z, A)$ would increase) for any suboptimal jamming strategy. We shall show that $G(P_X^*, P_Z, A^*) \geq G(P_X^*, P_Z^*, A^*)$ as follows. $G(P_X^*, P_Z, A^*)$ equals

$$\min_{\{f_{xy}\}} \sum_{xy} f_{xy} \ln (f_{xy}/(f_x f_y))$$

subject to

$$f_x = p_x \qquad \text{for all } x$$

$$f_y = p_y \qquad \text{for all } y$$

and

$$\sum_{xy} f_{xy} \ln p^*_{xy} \geq \sum_{xy} p_{xy} \ln p^*_{xy}$$

Adding

$$-\sum_x p_x \ln p^*_x - \sum_y p_y \ln p^*_y$$

to both sides of the inequality constraint, and using the fact that $f_x = p_x$ and $f_y = p_y$, the inequality constraint becomes

$$\sum_{xy} f_{xy} \ln (p^*_{xy}/(p^*_x p^*_y)) \geq \sum_{xy} p_{xy} \ln (p^*_{xy}/(p^*_x p^*_y))$$

or equivalently

$$\sum_{xy} f_x f_{y|x} \ln (p^*_{y|x} / p^*_y) \geq \sum_{xy} p^*_x p_{y|x} \ln (p^*_{y|x} / p^*_y) \quad 4.3.2$$

On the other hand,

$$\sum_{xy} f_{xy} \ln (f_{xy} / (f_x f_y))$$

$$= \sum_{xy} f_x f_{y|x} \ln (f_{y|x} / f_y) \quad 4.3.3a$$

$$\geq \sum_{xy} f_x f_{y|x} \ln (p^*_{y|x} / p^*_y) \quad b$$

with 4.3.3b following from

$$\sum_{xy} f_{xy} \{ \ln (p^*_{xy} / (p^*_x p^*_y)) - \ln (f_{xy} / (f_x f_y)) \}$$

$$= \sum_{xy} f_{xy} \ln (p^*_{xy} / f_{xy})$$

$$\leq \sum_{xy} f_{xy} (p^*_{xy} / f_{xy} - 1)$$

$$= 0$$

Combining (4.3.2) and (4.3.3b), we have

$$\sum_{xy} f_x f_{y|x} \ln (f_{y|x} / f_y) \geq \sum_{xy} p^*_x p_{y|x} \ln (p^*_{y|x} / p^*_y)$$

It remains to show that

$$\sum_{xy} p^* p_{y|x} \ln (p^*_{y|x} / p^*_y) \geq \sum_{xy} p^*_x p^*_{y|x} \ln (p^*_{y|x} / p^*_y)$$

$$= G(P^*_X, P^*_Z, A^*)$$

Define Q as the set of $P_{y|x}$ such that

$$p_{y|x} = \sum_z p_{y|xz}\ p_z$$

for some $P_z \in \mathcal{Z}$. Since $p_{y|x}$ is linear in p_z, and that $\mathcal{Z}$ is a convex set, it follows that Q also is a convex set. Define

$$F(P_{y|x}) = \sum_{xy} p^*_x\ p_{y|x}\ \ln\ (p_{y|x}\ /p^*_y)$$

By definition, $P^*_{y|x}$ achieves the minimum of F. For all $P^+_{y|x} \in Q$, define $P'_{y|x}$ such that

$$p'_{y|x} = (1-\epsilon)\ p^*_{y|x} + \epsilon\ p^+_{y|x}$$

$P'_{y|x} \in Q$ since it is a linear combination of $P^*_{y|x}$ and $P^+_{y|x}$. Now

$$F(P'_{y|x})$$

$$= F((1-\epsilon)p^*_{y|x} + \epsilon p^+_{y|x})$$

$$= F(p^*_{y|x} + \epsilon (p^+_{y|x} - p^*_{y|x}))$$

$$= F(p^*_{y|x}) + \sum_{xy} \left[\frac{\partial}{\partial p_{y|x}} F(P_{y|x}) \right]_{p_{y|x} = p^*_{y|x}} \cdot \epsilon (p^+_{y|x} - p^*_{y|x})$$

$$= F(p^*_{y|x}) + \epsilon \sum_{xy} [\ p^*_x \ln (p^*_{y|x}/p^*_y) + p^*_x\ ]\ (p^+_{y|x} - p^*_{y|x})$$

for small $\epsilon > 0$. Since

$$F(P'_{y|x}) \geq F(P^*_{y|x})$$

we have

$$\sum_{xy} \quad p^*_x \ln (p^*_{y/x} / p^*_y) + p^*_x) (p^+_{y/x} - p^*_{y/x}) \geq 0$$

which reduces to

$$\sum_{xy} \quad p^*_x \, p^+_{y/x} \ln (p^*_{y/x} / p^*_y) \geq \sum_{xy} p^*_x \, p^*_{y/x} \ln (p^*_{y/x} / p^*_y)$$

This completes the proof for 4.3.3b as well as theorems 4.4.3 and 4.4.1.

Q.E.D.

4.5 Multiple accessing with incomplete codebook knowledge and jamming

The extension of the results of the previous section to the multiple access environment is conceptually straight forward. The coding theorem in this section can be proved rigorously. The proof is omitted since it gives no new insight and is very complex in details and notations.

The asynchronous M-user multiple access channel with incomplete codebook knowledge and jamming is modeled as follows. Let there be M asynchronous sources sending independent messages. The user i, $1 \leq i \leq M$, uses a codebook C_i containing 2^{nR_i} equiprobably used codewords $x_i^n(j)$, $1 \leq j \leq 2^{nR_i}$. We assume that there is an (M+1)-th source, which tries to jam the M users. For the M users and the jammer, we impose the following constraints

$$b_i(x_i^n) = \sum_{k=1}^{n} b_i(x_{ik}) \leq n B_i \qquad 1 \leq i \leq M+1$$

We assume that the receiver m, $1 \leq m \leq M$, is assigned to obtain an estimate for the message of user m. In the process, the receiver m may have to estimate the messages of the other users. The receiver m, besides knowing the codebook of the user m, also knows some codebooks of the other M-1 users. The set of indices of C_i's that are known to the receiver m is defined as I_m. The receiver m performs joint decoding for the messages of the users indexed by I_m, while assuming a metric which takes into account the effect of the channel and the users indexed by $\bar{I}_m = \{1,2,\ldots M\} - I_m$. We say that $(R_1,\ldots,R_M)$ is in the capacity

region if there exists codebooks $C_1,\ldots,C_M$ with small error probability at each receiver.

We have seen in the previous section that by imposing a specific structure on the decoder, the signal of the jammer has the effect of a memoryless noise with stationary statistics. Thus for the receiver m with $|I_m|=L$, the problem of decoding the L messages is the same as the L-user multiple access channel with complete codebook knowledge, with the other M-L users plus the jammer treated as memoryless noise with stationary statistics. It can be shown [17] by random coding that for each

$$P(x_1,\ldots,x_{M+1},y) = P_1(x_1)\ldots P_{M+1}(x_{M+1})\, P(y/x_1,\ldots x_{M+1})$$

reliable communication is achievable at the receiver m if

$$\sum_{i\in\Omega_m} R_i < I(\{X_i\}_{i\in\Omega_m};Y/\{X_i\}_{i\in\bar{\Omega}_m})$$

for all $\Omega_m \in I_m$ with $m\in I_m$, and $\bar{\Omega}_m=I_m-\Omega_m$. Reliable communication is achieved for the system if the above inequalities are satisfied for all $1\leq m\leq M$. The region defined by the above inequalities (with $R_i>0$ for all i) is denoted as $\tilde{\mathcal{R}}$. Consequently, the achievable rate region $\mathcal{R}$ is given by

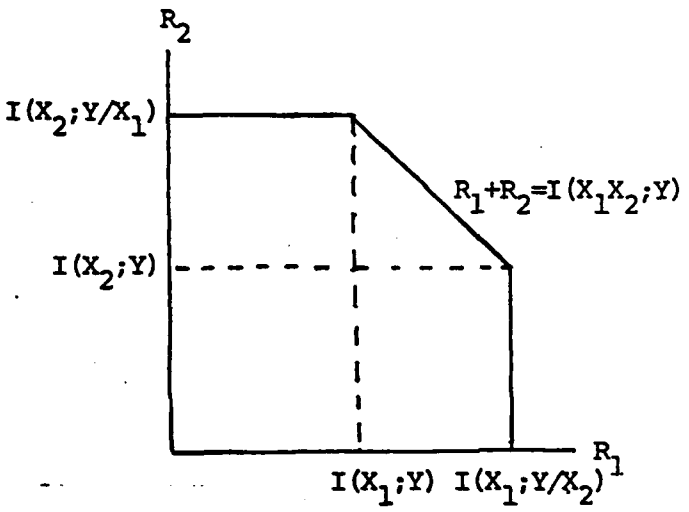
$$\mathcal{R} = \bigcap_{P(X_{M+1})\in\mathcal{X}_{M+1}} \; \bigcup_{P(X_i)\in\mathcal{X}_i,\, 1\leq i\leq M} \tilde{\mathcal{R}}$$

The convex hull is absent in the above characterization because the users are asynchronous. Furthermore, the above union and intersection can be reversed in order without changing $\tilde{\mathcal{R}}$. This reversibility results from a convexity argument and the saddle

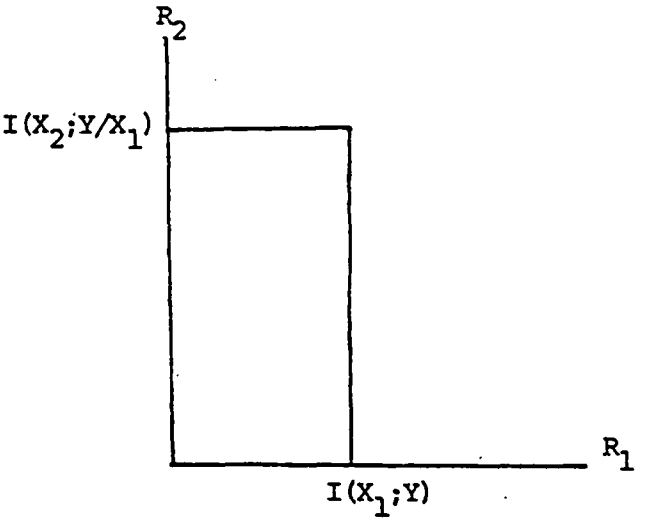
point theorem.

The converse, namely that the system has bit error probabilities bounded above zero for any code outside the region $\mathcal{R}$, can be proved by standard techniques [16] and the techniques used in chapter 2.

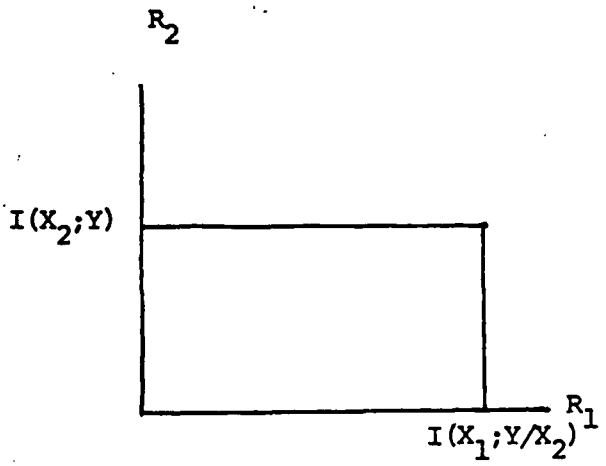
Figure 4.5.1 illustrates the region $\tilde{\mathcal{R}}$ for M=2 N and various degree of codebook knowledge. Obviously, more codebook knowledge results in a larger $\tilde{\mathcal{R}}$.



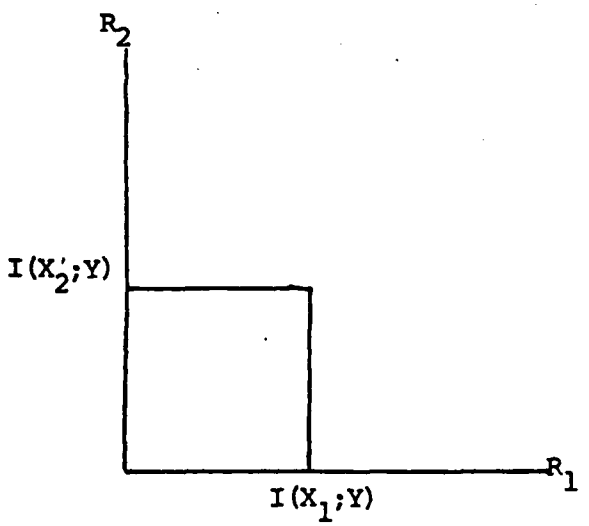
$I_1 = \{1,2\}$
$I_2 = \{1,2\}$



$I_1 = \{1\}$
$I_2 = \{1,2\}$



$I_1 = \{1,2\}$
$I_2 = \{2\}$



$I_1 = \{1\}$
$I_2 = \{2\}$

Figure 4.5.1 $\tilde{\mathcal{R}}$ for various degree of codebook knowledge

Appendix 4.1 Upper bound for probability of atypicality

We want to show that for all $\epsilon' > 0$, there exists some n_o such that

$$P_{\mathbb{C}}(Y^n \notin T^n_{\epsilon,y}) < \epsilon'$$

if $n > n_o$.

Proof

$$P_{\mathbb{C}}(Y^n \notin T^n_{\epsilon,y})$$

$$= P(Y^n \notin T^n_{\epsilon,y} \,/ X^n(j) \in T^n_{\epsilon,x})$$

$$= P(Y^n \notin T^n_{\epsilon,y}, X^n(j) \in T^n_{\epsilon,x}) \,/\, P(X^n(j) \in T^n_{\epsilon,x})$$

$$\leq P(Y^n \notin T^n_{\epsilon,y}, X^n(j) \in T^n_{\epsilon,x}) / (1-\epsilon) \qquad \text{provided } n > \text{some } n'_o$$

$$= 1/(1-\epsilon) \sum_{(x^n,y^n): x^n \in T^n_{\epsilon,x},\, y^n \notin T^n_{\epsilon,y}} P(x^n y^n)$$

$$\leq 1/(1-\epsilon) \sum_{y^n \notin T^n_{\epsilon,y}} P(y^n) \qquad \text{provided } n > \text{some } n''_o$$

$$\leq \epsilon/(1-\epsilon)$$

$$= \epsilon'$$

Hence by choosing $\epsilon = \epsilon'/(1+\epsilon')$, we have

$$P_{\mathbb{C}}(Y^n \notin T_{\epsilon,y}) \leq \epsilon'$$

if $n > \max(n'_o, n''_o) \triangleq n_o$

The proof for the fact that for all $\epsilon'>0$, there exists some n_0 such that

$$P_{\mathcal{C}}((X^n(j),Y^n) \notin T^n_{\epsilon,+}) \leq \epsilon'$$

if $n>n_0$ follows the same argument.

Appendix 4.2 Upper bounds for $P_{\mathcal{C}}(x^n(j))$ and $P_{\mathcal{C}}(y^n)$

We want to upper bound $P_{\mathcal{C}}(x^n(j))$ and $P_{\mathcal{C}}(y^n)$. Now

$$P_{\mathcal{C}}(x^n(j))$$

$$= P(x^n / x^n \in T^n_{\epsilon,x})$$

$$= P(x^n \text{ and } x^n \in T^n_{\epsilon,x}) \; / \; P(x^n \in T^n_{\epsilon,x})$$

$$\leq 1/(1-\epsilon) \quad P(x^n \text{ and } x^n \in T^n_{\epsilon,x})$$

$$= 1/(1-\epsilon) \quad \prod_x \quad p_x^{n_x}$$

$$\leq 1/(1-\epsilon) \quad \prod_x \quad p_x^{n(p_x - \epsilon)}$$

$$\leq 1/(1-\epsilon) \quad \exp\left(-n(H(X) + \epsilon \sum_x \ln p_x)\right)$$

$$\leq \exp\left(-n(H(X) - \epsilon')\right)$$

in which ϵ' is small, thus giving the desired upper bond.

The proof for the bound

$$P_{\mathcal{C}}(y^n) \leq \exp\left(-n(H(Y) - \epsilon')\right)$$

follows the same argument.

Chapter 5 Multiple Accessing for the OR Channel

One convenient way of communication over the satellite channel[4] or the optical channel[19] is for each user to signal in pulses. We shall model such channels as the OR channel defined as follows. The channel alphabet is $X=\{0,1\}$. An idle user transmits the idle symbol 0. The channel output alphabet is $Y=\{0,1\}$. The channel transitions are given by $y=0$ if all $x_i=0$, otherwise $y=1$. The slotting of the channel may not be synchronized among the users. However, the channel can be modeled as symbol synchronous by limiting the temporal resolution and using a discrete estimate for the position of the pulse.

This chapter investigates the efficiency of the OR channel for multiple accessing. The communication system has a large number of users, and only a small fraction is transmitting at a time. We shall assume that the system is asynchronous.

This chapter is divided as follows. Section 5.1 evaluates the channel capacity and the cutoff rate of this channel. For the case with joint decoding, the sum-capacity of one (which is also the sum-capacity of the synchronous channel) can be achieved. For the case without joint decoding, the sum-capacity and the sum-cutoff rate both equal ln 2 (or .69). Section 5.2 analyses a convolutional encoding and decoding scheme. The sum-cutoff rate is used to obtain a convenient upper bound for the error probability as a function of the constraint length of the convolutional code and the sum-throughput of the users. We define

complexity as n, the number of possibly correct states in the trellis in steady state. Section 5.2 also shows , with an independence assumption, that P(n) decreases exponentially in n for large n. Hence the probability of buffer overflow as n exceeds the size of a modestly large buffer (that registers the set of possible states) is almost negligible. This decoding algorithm shows that a sum-throughput close to the sum-cutoff rate of 0.69 is achievable in practice.

5.1 Capacity and cutoff rate.

For the asynchronous multiple access channel with joint decoding, the capacity region is given by

$$\mathcal{R} = \bigcup_{P_1(X_1),\ldots,P_M(X_M)} \tilde{\mathcal{R}}$$

$$:P(x_1 \ldots x_M y)=P_1(x_1)\ldots P_M(x_M).P(y/x_1 \ldots x_M)$$

in which $(R_1,\ldots,R_M) \in \tilde{\mathcal{R}}$ if

$$\sum_{i\in\Omega} R_i < I(\{X_i\}_{i\in\Omega}; Y/\{X_i\}_{i\in\bar{\Omega}})$$

for all

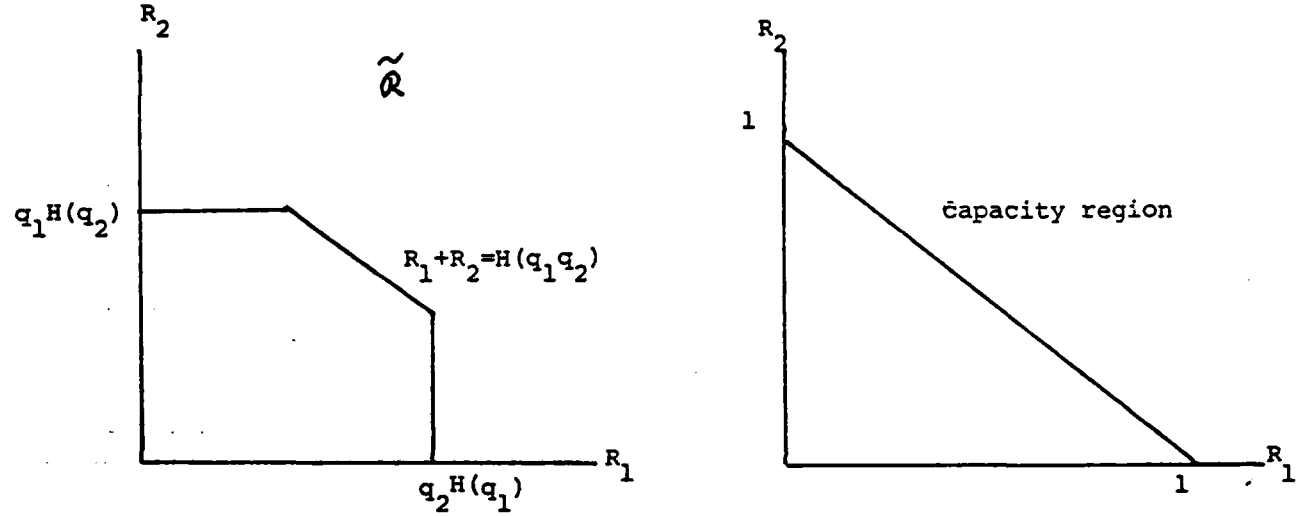
$$\Omega \subseteq \{1,\ldots,M\},\ \Omega \neq \phi$$

$$\bar{\Omega} = \{1,\ldots,M\} - \Omega$$

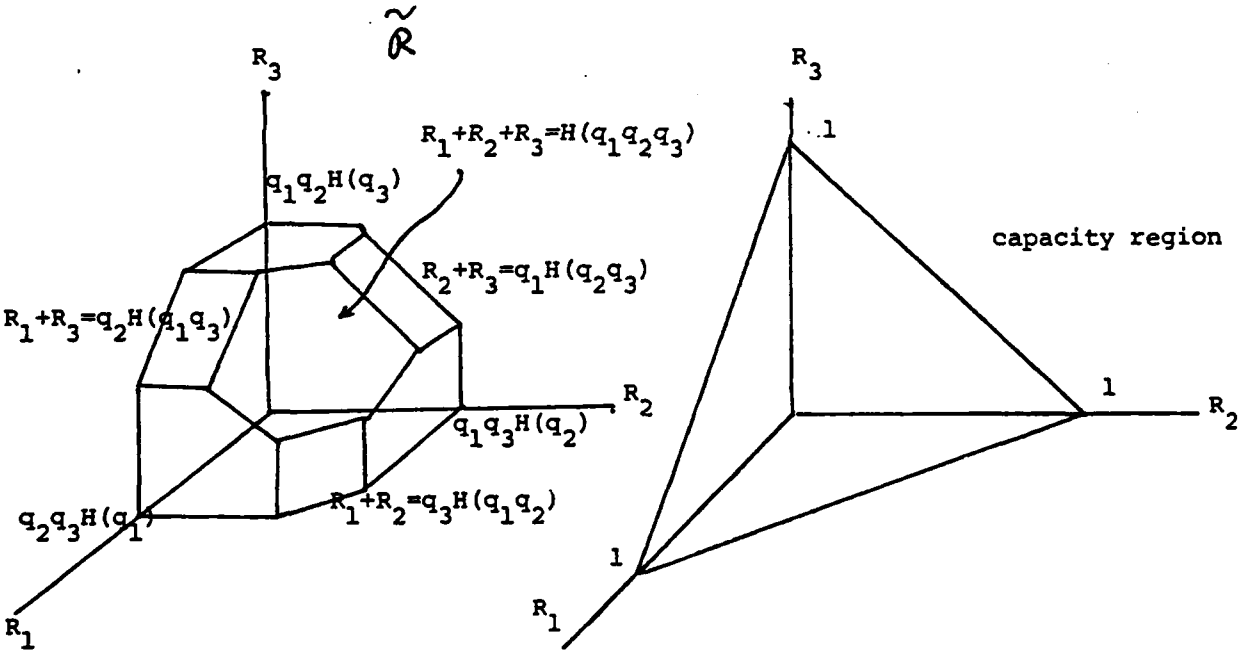
For $1\leq i\leq M$, let $P_i(x_i=0)=q_i$ and $P_i(x_i=1)=1-q_i$. It can be readily shown that

$$I(\{X_i\}_{i\in\Omega}; Y/\{X_i\}_{i\in\bar{\Omega}}) = \left(\prod_{i\in\bar{\Omega}} q_i\right) H\left(\prod_{i\in\Omega} q_i\right)$$

The right hand side of the above equation can be interpreted as the product of the probability that the users in the set $\bar{\Omega}$ do not put a pulse in a position with the binary entropy of the output given that the users in the set $\bar{\Omega}$ do not put a pulse in the position. The region $\tilde{\mathcal{R}}$ (for a specific set of $P_1(X_1),\ldots,P_M(X_M)$) is shown in figure 5.1.1 for the cases of M=2 and M=3. It is noteworthy that many $\tilde{\mathcal{R}}$ are required to form the capacity region shown in figure 5.1.1.



M = 2



M = 3

Figure 5.1.1 The asynchronous OR channel with joint decoding

It can be readily shown that the point $(1/M,1/M,\ldots,1/M)$ is in the $\mathcal{R}$ with $q_i=(1/2)^{1/M}$ for all i. Thus a sum-throughput of 1 is achievable for the asynchronous OR channel with joint decoding. Obviously, the sum-throughput cannot exceed 1, the maximum entropy per letter for the output alphabet.

For the case without joint decoding, the capacity region equals

$$\mathcal{R} = \bigcup_{P_1(X_1),\ldots,P_M(X_M)} \tilde{\mathcal{R}}$$

$$:P(x_1\ldots x_M y)=P_1(x_1)\ldots P_M(x_M)P(y/x_1\ldots x_M)$$

in which $(R_1,\ldots,R_M) \in \tilde{\mathcal{R}}$ if

$$R_i < I(X_i;Y)$$

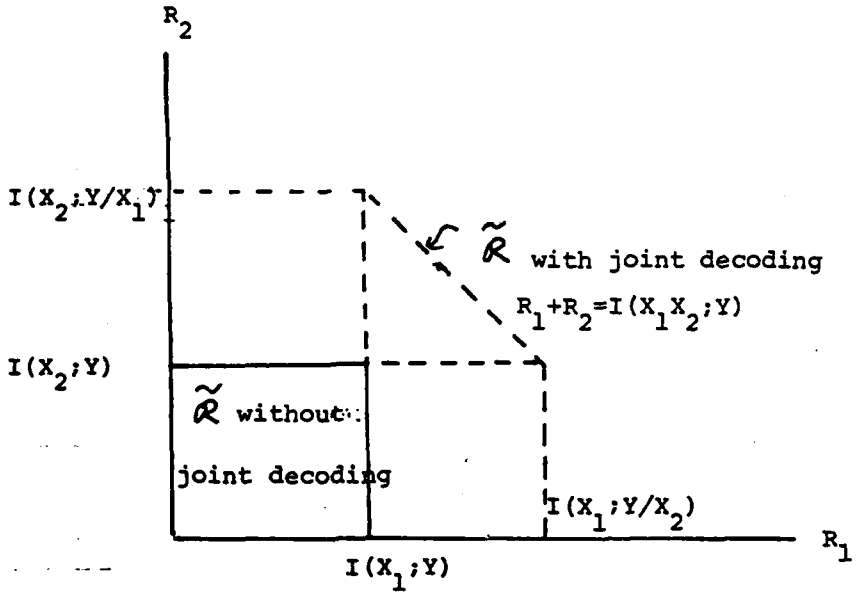
for all i such that $1\leq i\leq M$. The capacity region for the asynchronous 2-user OR channel without joint decoding is illustrated in figure 5.1.2. It can be readily shown that

$$I(X_i;Y) = H\left(\prod_{j=1}^{M} q_j\right) - q_i\, H\left(\prod_{j=1}^{M} q_j/q_i\right)$$

in which $P(x_i=0)=q_i$. Consider R_T, the sum of rates for the M users with $q_i=q$ for all i.

$$R_T = M\, H(q^M) - M\, q\, H(q^{M-1})$$

Substituting $q^M = r$ and $M=1/x$ in the above expression gives



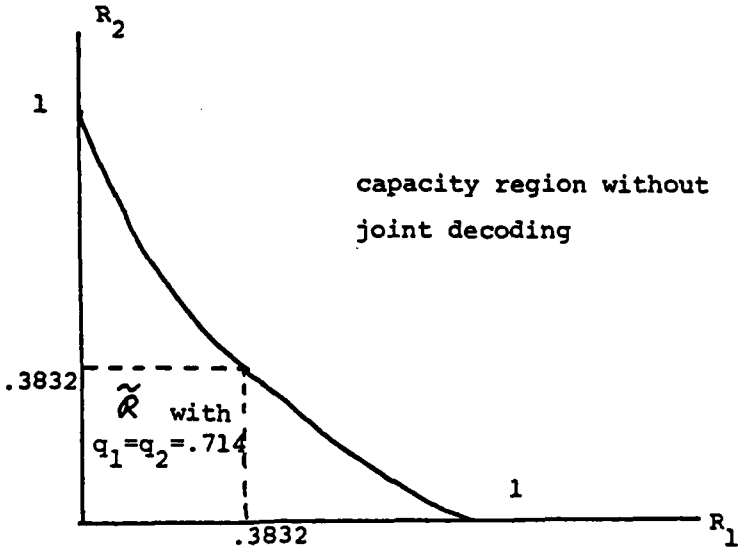


Figure 5.1.2 The asynchronous OR channel without joint decoding

$$R_T(r,x)$$

$$= [-xr \ln r - (1-r) \ln(1-r) + r^{x}(1-r.r^{-x}) \ln(1-r.r^{-x})]/x \text{ nats}$$

For large M, the above expression can be evaluated using L'Hospital's rule with x going to zero, subsequently giving

$$R_T(r) \triangleq \lim_{x \to 0} R_T(r,x) = \ln r \ln (1-r) \quad \text{nats}$$

The maximum of $R_T(r)$ over r occurs at r=1/2 when

$$R_T(r) = \ln 1/2 \ \ln 1/2 \ \text{nats} = \ln 2 \text{ bits} = .69 \text{ bits}$$

Thus the sum-capacity of the asynchronous OR channel without joint decoding is .69.

The sum-cutoff rate for the asynchronous OR channel treating the other users as memoryless noise happens to be the same as the capacity of the channel. In the evaluation of the cutoff rate, the modulation symbols have to be defined carefully. The sum-cutoff rate for the pulse position modulation is .69 while that for on-off keying is .59. The difference between these two sum-cutoff rate is due to the difference of the modulation symbols. For pulse position modulation, a symbol is a frame of N slots with a pulse in one of the slots. The set of modulation symbols consists of N different pulse positions. On the other hand, the set of modulation symbols of on-off keying is the set {0,1} representing the absence or presence of a pulse in a slot.

For pulse position modulation, we shall approximate the other users as memoryless (slot-wise) noise in the evaluation of

the cutoff rate for each user. For pulse position modulation, the cutoff rate per frame for user i is

$$R_0^i = -\ln \left(\sum_y \left(\sum_x Q(x) \sqrt{P(y/x)} \right)^2 \right) \qquad \text{nats/PPM symbol}$$

in which $Q(x)=1/N$ and y is an N-dimensional binary vector. Let h be the probability that a slot contains a pulse due to the users other than user i. It can be easily shown that for pulse position modulation,

$$h = 1 - (1 - 1/N)^{M-1} \approx 1 - e^{M/N} = 1 - e^{-T}$$

for large M and N; and $T=M/N$. The channel transition probability as viewed by user i (with the memoryless approximation) is given by

$$P(y/x_i) = (1-h)^{N-k} h^{k-1}$$

for those y with k pulses, one of which falling into the pulse position of the code symbol x_i. $P(y/x_i)=0$ if there is no pulse in the position of y that corresponds to the pulse position of x_i. Thus

$$R_0^i = -\ln \left(\sum_{k=1}^{N} {}_NC_k \, (k/N \;\; (1-h)^{N-k} h^{k-1})^2 \right)$$

$$= -\ln \left((1-h)^N/(hN^2) \sum_{k=1}^{N} k^2 \, {}_NC_k \, [h/(1-h)]^k \right)$$

The summation above can be readily shown to be equal to

$$N(N-1) \, (h/[1-h])^2 \;\; (1/[1-h])^{N-2} + N \, (h/[1-h]) \, (1/[1-h])^{N-1}$$

Assuming large N (say > 10), the second term in the above expression is negligible compared with the first term. Further algebraic manipulation gives

$$R_o^i = -\ln h \approx -\ln (1-e^{-T}) \quad \text{nats/PPM symbol}$$

The sum-cutoff rate per slot is defined as

$$R_T = -M/N \; R_o^i = -T \ln (1-e^{-T})$$

which is maximized when T=ln 2 with

$$R_T = -\ln 2 \ln (1/2) \quad \text{nats/slot} = \ln 2 \quad \text{bits/slot}$$

Thus, the sum-cutoff rate is the same as the sum-capacity, with the memoryless approximation.

For on-off keying, the channel transition probabilities for user i is given in figure 5.1.3 (assuming $q_i = q$ for all users). The sum-cutoff rate is then

$$R_o = -M \ln \left(\sum_y \left(\sum_x Q(x) \sqrt{P(y/x)} \right)^2 \right) \quad \text{nats/slot}$$

$$= -M \ln (1 - 2q + 2q^2 + 2q(1-q)(1-q^{M-1})^{1/2})$$

Substituting $q^M = r$ and $x = 1/M$ gives

$$R_o(r,x) = -\ln (1 - 2r^x + 2r^{2x} + 2r^x(1-r^x)(1-r.r^{-x})^{1/2}) / x$$

Applying L'Hospital's rule for large M gives

$$R_T(r) = \lim_{x \to 0} R_T(r,x) = 2 (1 - \sqrt{1-r}) \log_2 r \quad \text{bits/slot}$$

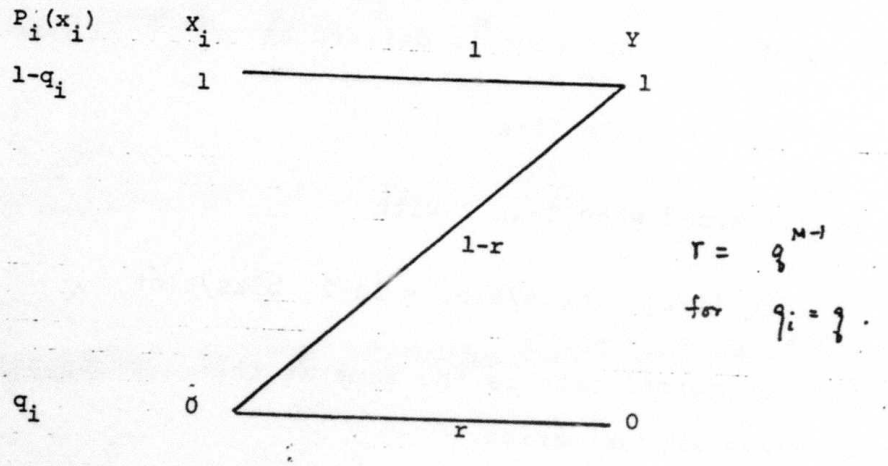


Figure 5.1.3 Channel transition diagram for on-off keying

with a maximum at r=.421 when R = .597, which is less than that for pulse position modulation channel. We are unable to give convincing reasons to explain the difference.

5.2 Convolutional codes for the multiple access OR channel

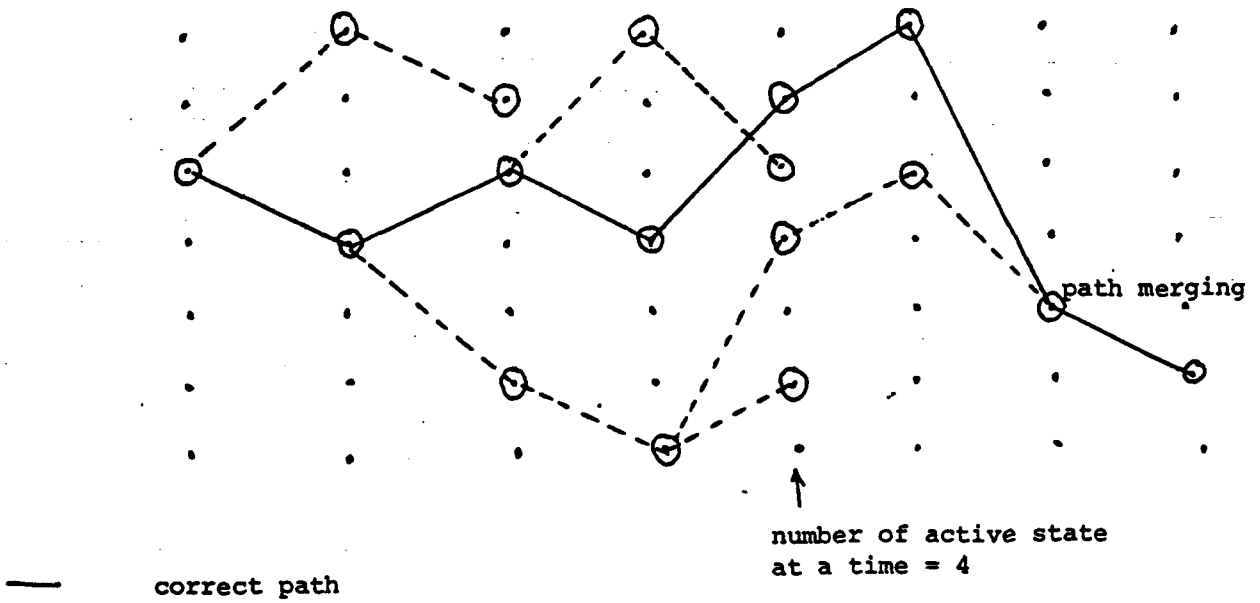
This section analyses the throughput, error probability and decoding complexity of a specific coding and decoding scheme. Each user is assigned a binary input N-ary output convolutional encoder as shown in figure 1.2.1. The tap gains a_i's are integers chosen in the set $\{0,1,2,\ldots,N-1\}$. Decoding is performed as follows. We assume that the transmitter-receiver pair is code-synchronized. Thus the code sequence sent by the transmitter traverses a path in the binary trellis (the solid line in figure 5.2.1). The pulses of the other users may turn on branches from the correct path. The state at the end of a turned-on branch is stored in a buffer, together with the path that leads to the state. We call such a state an active state. A path may be further extended if there is a channel pulse at the pulse position associated with the two branches from the active state.

Decoding errors may result when an incorrect path merges with the correct path. The ensemble average of the expected number of bit errors for an error event starting at a given time is upper bounded in [10] by

$$P_b < 2^{-k R_0^i} / [1 - 2^{(R_0^i - 1)}]^2$$

$$= (1 - e^{-T})^k / (2 e^{-T} - 1)^2$$

in which k is the constraint length and $R_0^i = -\log_2(1-e^{-T})$ bits/PPM symbol which is derived in section 5.1 using the memoryless noise assumption. (Note: the rate of the encoder is 1



correct path

incorrect path

active state

Figure 5.2.1 The decoding algorithm

bit/PPM symbol because the tree is binary.) This upper bound is plotted in figure 5.2.2 as a function of the sum-throughput T for various values of k. From the plot, it is necessary to use a large k to make P_b small.

The complexity of decoding is measured by the random variable n, the number of active states in steady state. Assume that the decoder has a buffer that can register at most L active states. The buffer overflows when the number of active state exceeds L, consequently the correct path may not be extended. Thus a decoding failure may also occur when the buffer overflows. We shall demonstrate that P(n) decreases exponentially in n for large n, using an independence assumption. Hence the probability of buffer overflow can be made very small using a modestly large buffer. In contrast, the probability of buffer overflow for sequential decoding on typical channel decreases like a power of 1/L, which is much worse than the exponential decrease in L for our case.

Let n_j be the number of active states at time j in the trellis. The number of active states for the trellis is upper bounded by the number of active states when the trellis is considered as a tree, by double counting the active states resulting from two merging paths. For large k, path merging is highly unlikely, hence treating the trellis as a tree would give a tight bound. Initially, there is only one active state, that is $n_1=1$. There is always an active state propagated by the correct path at each time. Without loss of generality, this state

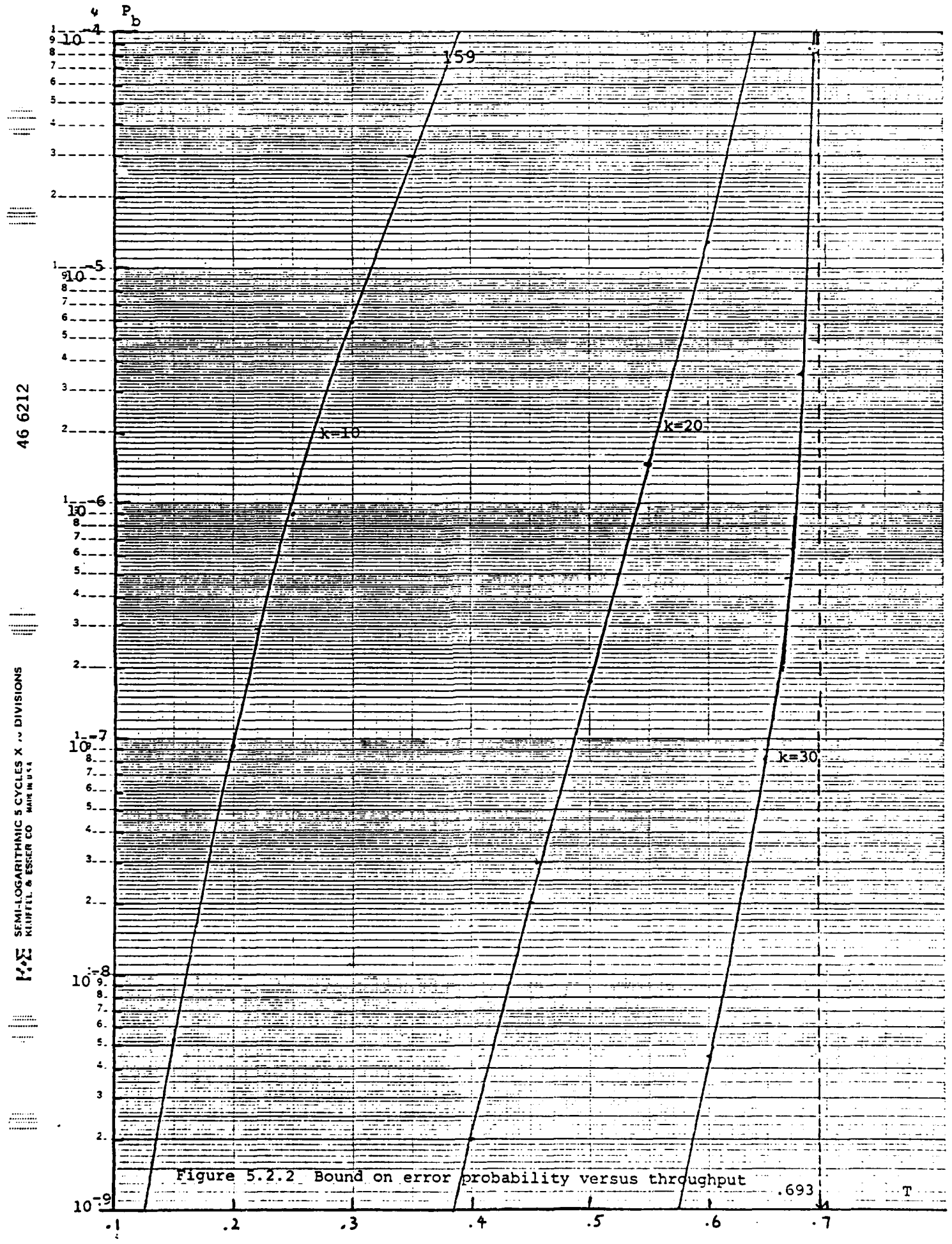


Figure 5.2.2 Bound on error probability versus throughput

is labeled 1. The other active states are labeled i, $2 \leq i \leq n_j$. Let $e_{i,j}$ be the number of states at time j+1 that are activated by the active state i at time j. Let p be the probability that a pulse due to the other users falls in any slot. (For M users and N-ary PPM symbols, recall that $p=1-e^{-T}$, where T is the channel throughput M/N (bits/slot)). The $e_{i,j}$'s for given j are not independent since the $e_{i,j}$'s depend on the received channel sequence. However, the $e_{i,j}$'s would be very weakly coupled for large M and N, in the sense that knowing one $e_{i,j}$ tells little about the other $e_{i,j}$'s. The $e_{i,j}$'s have probability distributions.

$$P(e_{1,j} = 1) = 1-p$$

$$P(e_{1,j} = 2) = p$$

and for $2 \leq i \leq n_j$

$$P(e_{i,j} = 0) = (1-p)^2$$

$$P(e_{i,j} = 1) = 2p\,(1-p)$$

and

$$P(e_{i,j} = 2) = p^2$$

We shall drop the superfluous subscript j henceforth. The probability generating functions for $P(e_i)$ are defined as

$$V_1(s) = \sum_{k=0}^{\infty} P(e_1 = k)\, s^k = (1-p)\, s + p\, s^2$$

and

$$V_i(s) = \sum_{k=0}^{\infty} P(e_i = k)\, s^k$$

$$= (1-p)^2 + 2p(1-p)\, s + p^2 s^2$$

$$= [sp + (1-p)]^2$$

for $2 \leq i \leq M$.

The sequence of random variables $n_1, n_2, \ldots, n_j, \ldots$ constitutes a discrete time branching process [15]. Furthermore

$$n_{j+1} = \sum_{i=1}^{n_j} e_i$$

with probability generating function

$$S_{j+1}(s)$$

$$= \sum_{k=0}^{\infty} P(n_{j+1} = k)\, s^k$$

$$= \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} P(n_{j+1} = k / n_j = m)\, P(n_j = m)\, s^k$$

$$= \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} P(n_j = m)\, P(e_1 + \ldots + e_m = k)\, s^k$$

$$= \sum_{m=0}^{\infty} P(n_j = m) \sum_{k=0}^{\infty} P(e_1 + \ldots + e_m = k)\, s^k$$

$$= \sum_{m=0}^{\infty} P(n_j = m) \prod_{i=1}^{m} V_i(s)$$

$$= \sum_{m=0}^{\infty} P(n_j = m)\, [(1-p)\, s + p\, s^2]\, [sp + (1-p)]^{2(m-1)}$$

$$= \quad s/[sp + (1-p)] \quad \sum_{m=0}^{\infty} P(n_j = m) \, [sp + (1-p)]^{2m}$$

$$= \quad s/[sp + (1-p)] \quad S_j([sp + (1-p)]^2)$$

in which the product $\prod_{i=0}^{M} V_i(s)$ results from the assumption that the e_i's are independent random variables, and the fact that the probability generating function of a sum of independent random variables equals the product of the probability generating functions of the random variables. In steady state, dropping the subscripts j and j+1 gives

$$S(s) = s/[sp + (1-p)] \quad S([sp + (1-p)]^2)$$

If $S(s) < \infty$ for some $s>1$, $P(n)$ must decreases faster than s^{-n} for large n. Therefore, we are interested in the value of S(s) for s>1. For the sake of providing insights about the function S(s), we give a recursive algorithm to compute S(s). Let

$$s_k = [s_{k+1} p + (1-p)]^2$$

Thus

$$s_{k+1} = (\sqrt{s_k} - (1-p)) / p$$

The plot of s_{k+1} versus s_k is given in figure 5.2.3. Thus, we have the recursive relation

$$S(s_{k+1}) = s_{k+1} / \sqrt{s_k} \quad S(s_k)$$

From figure 5.2.3, $s_{k+1} > s_k$ if $1 < s < ((1-p)/p)^2$. The recursive relationship gives the value of S for progressively

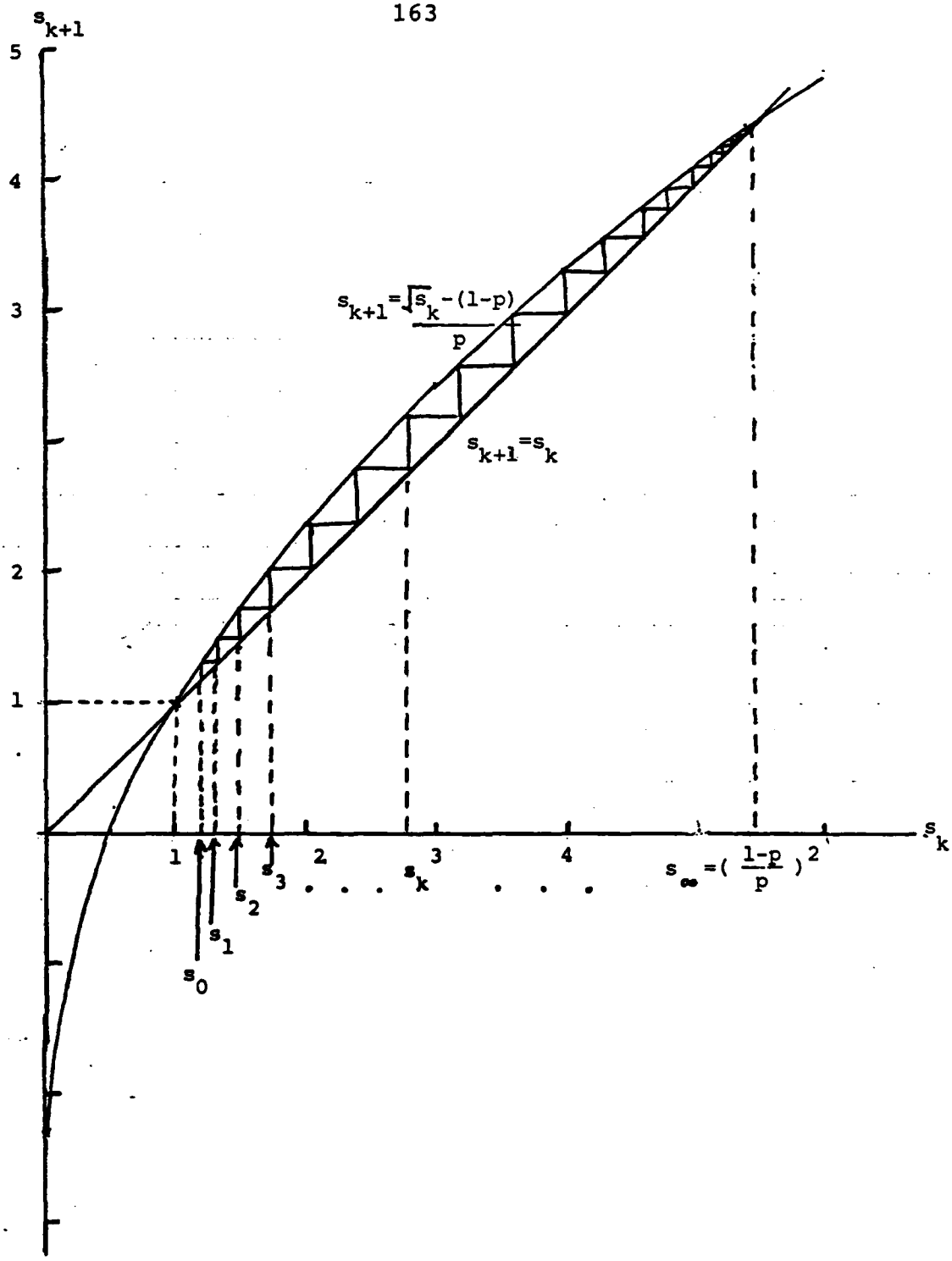


Figure 5.2.3 Computing $S(s_k)$ recursively

larger s, which inches towards the value $((1-p)/p)^2$ as k increases. Thus, if we start with $s_0=1+\epsilon$ with known $S(s_0)$, we can compute all $S(s_k)$, k=1,2,.... By varying the value of ϵ (which is assumed small), all values of S(s) in the open interval $(1,((1-p)/p)^2)$ can be computed. From the recursive relation, it is obvious that S(s) is finite for s in this open interval provided $S(1+\epsilon)$ is finite for some $\epsilon>0$, since any s in the interval can be reached in a finite number of steps of recursion. When s approaches $((1-p)/p)^2$, the increase of s_k for each step is small, while each successive value of S increases by a factor of $s_{k+1}/\sqrt{s_k}=(1-p)/p$. Therefore S becomes unbounded as s approaches $((1-p)/p)^2$. The remaining question is how to compute $S(1+\epsilon)$. Obviously S(1)=1. Differentiating the equation for S(s) with respect to s, and putting s=1 (so that $S'(1)=\bar{n}$), we have

$$\bar{n} = (1-p)/(1-2p)$$

Thus for small ϵ, we have

$$S(1+\epsilon) = 1 + ((1-p)/(1-2p))\,\epsilon$$

Differentiating twice with respect to s, we obtain

$$S''(1) = S'(1)\,(\,2p(1+p)\,)/\,(\,(1-2p)(1+2p)\,)$$

Hence, the variance of n is

$$\sigma_n^2 = \bar{n^2} - (\bar{n})^2 = (\,p(1-p)\,)/\,(\,(1-2p)^2(1+2p)\,)$$

The values of $\bar{n}$ and σ_n are plotted in figure 5.2.4 as a function of the sum-throughput T.

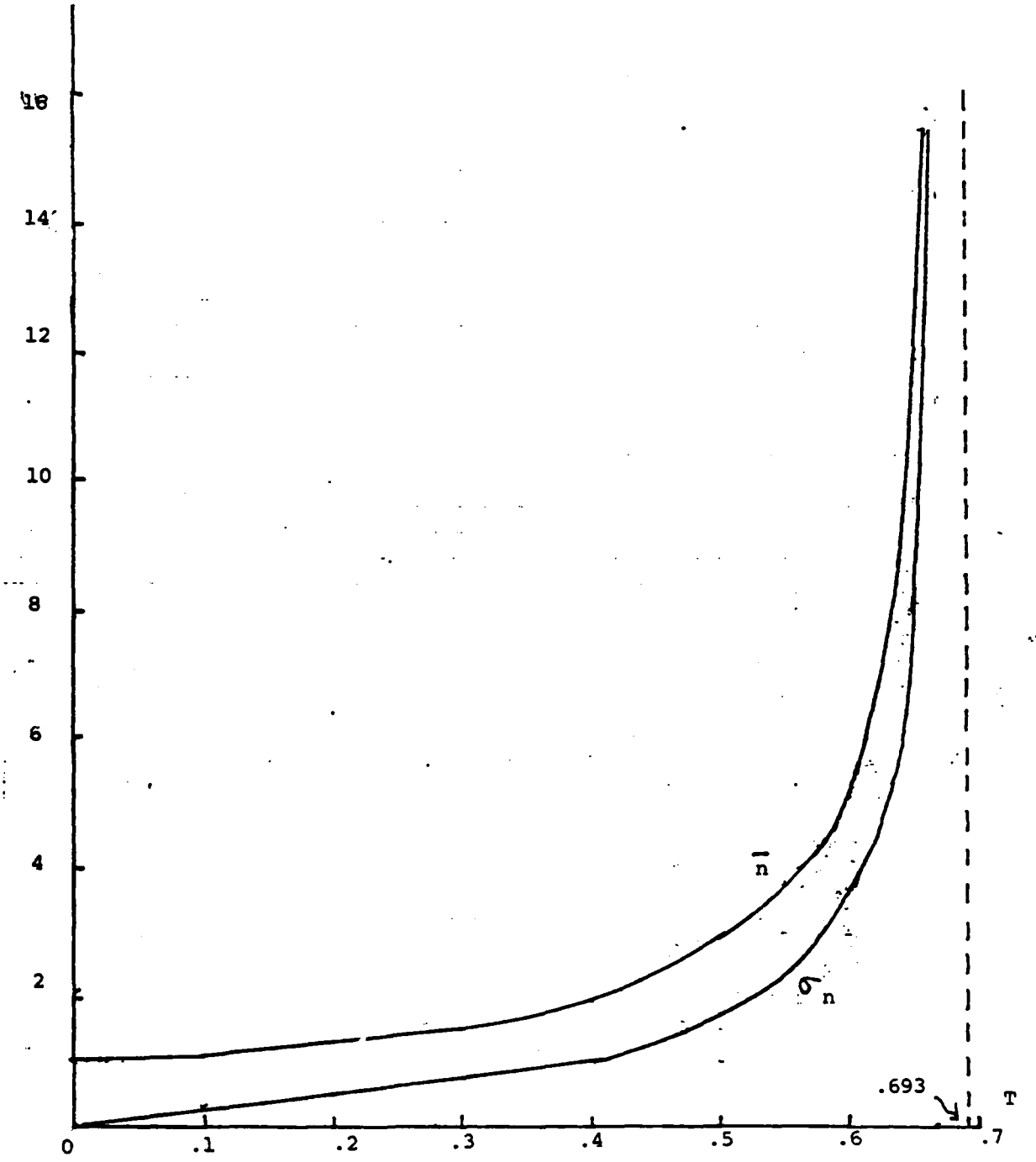


Figure 5.2.4 Mean and standard deviation of n versus throughput

Unfortunately, this algorithm is numerically unstable. Furthermore, we have not shown that $S(1+\epsilon)$ necessarily exists. The following upper bound on $S(s)$ would resolve these problems.

Theorem 5.2.1

$S_k(s) \leq (s^*-1)s/(s^*-s)$

for $s^*>s\geq 1$, $s^*=((1-p)/p)^2 > 1$

Proof (by induction)

$S_1(s) = s \leq (s^*-1)s/(s^*-s)$

for $s^*>s\geq 1$

Suppose

$S_k(s) \leq (s^*-1)s/(s^*-s)$

for $s^*>s\geq 1$, then

$$S_{k+1}(s)$$

$$= sp/[sp+(1-p)]\ S_k([sp+(1-p)]^2)$$

$$\leq sp/[sp+(1-p)]\ (s^*-1)[sp+(1-p)]^2/(s^*-[sp+(1-p)]^2)$$

$$= sp\ (s^*-1)\ (sp+(1-p))/[(s^*-s)(sp^2+(1-p^2)]$$

$$\leq (s^*-1)s/(s^*-s)$$

if $s^*>s$ and

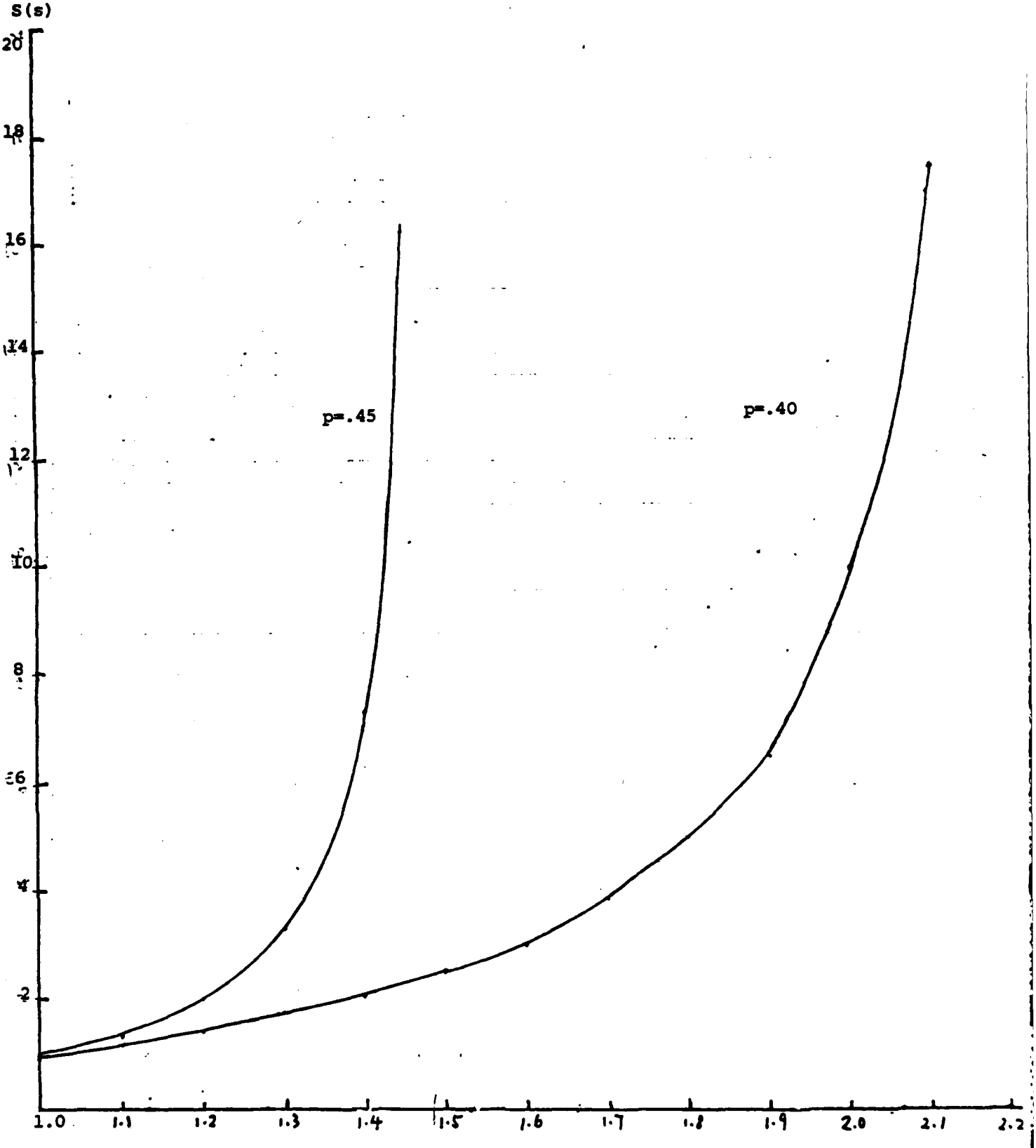


Figure 5.2.5 Probability generating functions for n

$p(sp+(1-p))/(sp^2+(1-p^2)) \leq 1 \quad (<=> \quad p \leq 1)$

Hence by induction, the theorem is true

Q.E.D.

Moreover, $S_k(s)$ is monotonically increasing in k, as stated in

Theorem 5.2.2

$S_{k+1}(s) \geq S_k(s)$

for $k \geq 1$ and $s^* > s \geq 1$

Proof (also by induction)

$S_2(s) - S_1(s)$

$= [s^2p + s(1-p)] - s$

≥ 0

for $s \geq 1$.

Suppose

$S_{k+1}(s) \geq S_k(s)$

for $s^* > s \geq 1$. Then

$S_{k+2}(s) - S_{k+1}(s)$

$= sp/[sp+(1-p)] \{S_{k+1}([sp+(1-p)]^2) - S_k([sp+(1-p)]^2)\}$

≥ 0

for $s^* > s \geq 1$. Hence

$S_{k+1}(s) \geq S_k(s)$

is true for all k by induction.

Q.E.D.

Since $S_k(s)$ is upper bounded and increases monotonically in k, $S_k(s)$ converges to an S(s), the steady state probability generating function. Thus, a steady state probability distribution P(n) exists, and the tail of P(n) decreases faster than $(p/(1-p))^{2n}$ for large n. The upper bound $(s^*-1)s/(s^*-s)$ is plotted in figure 5.2.5 for p=.4 and .45, when the sum-throughputs are .51 and .60 respectively.

The above scheme also achieves the sum capacity of .69. $S(1+\epsilon)$ is finite for some $\epsilon>0$ provided $((1-p)/p)^2 > 1$, or $p<1/2$. Therefore, reliable communication can be achieved if the sum-throughput T satisfies $p=1-e^{-T} < 1/2$, or equivalently $T<.69$.

Renouncing the independence assumption leads to a Pareto distribution [10] for P(n). The same phenomenon happens to the collision channel, which is analysed in chapter 7.

Chapter 6 Multiple Accessing for the Spread Spectrum Channel

This chapter considers the use of error correction codes on top of Psuedo-Noise (PN) sequence coding for code division multiple accessing of the asynchronous spread spectrum channel. The channel is found to have a maximum throughput of .72 and .36 based on the evaluation of channel capacity and cut-off rate respectively. More generally, these two values are derived for given n/N in which n is the length of the PN sequence and N the number of simultaneous users. It is found that to achieve the maximum throughput for a given bandwidth expansion, n should be small. This implies that coding schemes with short PN sequences and low rate codes are superior in terms of throughput or antijam capability. The extreme case of n=1 corresponds to using a very low rate code with no PN sequence coding. Convolutional codes are recommended and analysed for their error rate and decoding complexity.

This chapter is divided as follows. Section 6.1 describes the coding scheme and subsequently models the channel. A Gaussian approximation is used in the process. Section 6.2 then evaluates the total capacity and cut-off rate for the active users. For convolutional codes and sequential decoding, it is shown that the maximum throughput is (log.e)/4 = .36. Section 6.3 gives an analysis for the bit error probability and decoding complexity as a function of the constraint length and rate of the convolutional code, as well as the length of the PN sequence.

6.1 Modeling

Let N be the number of active users. Each user i, $1 \leq i \leq N$, is assigned an n bit PN sequence $\{c_{i,j}\}_{j=1}^{n}$, $c_{i,j} \in \{\pm 1\}$. A chip is defined as as an interval of length d. The PN sequence carrier is the function

$$c_i(t) = \sum_{j=1}^{n} c_{i,j}\, s(t-jd)$$

in which s(t) equals 1 if $0 \leq t \leq d$ and 0 elsewhere. Each user encodes the binary (0 or 1) data stream $\{\ldots u_{i,-1}\, u_{i,0}\, u_{i,1} \ldots\}$ into the antipodal (1 or -1) code stream $\{\ldots x_{i,-1}\, x_{i,0}\, x_{i,1} \ldots\}$. The rate of encoding is r (≤ 1). A coding scheme using convolutional code is shown in figure 1.1.3. The signal sent by each user is

$$x_i(t) = \sqrt{P_i} \sum_{k=-\infty}^{\infty} x_{i,k}\, c_i(t-knd+D_i)$$

in which P_i is the power for transmitter i, and D_i is the delay for user i, which is randomly distributed in [0,nd]. The channel output is

$$y(t) = \sum_{i=1}^{N} x_i(t)$$

The receiver i first of all acquires synchronization with transmitter i by estimating D_i, which we assume can be estimated accurately by some means. Demodulation is performed by match filtering and the values

$$y_{i,k} = \left(n \sum_{\substack{j=1 \\ j \neq i}}^{N} P_j\right)^{-1/2} / d \int_{-D_i+knd}^{-D_i+(k+1)nd} y(t)\, c_i(t-knd+D_i)\, dt$$

are obtained. The $y_{i,k}$'s are subsequently quantized. We shall assume either hard quantization or no quantization. The channel, as viewed by user i, can be characterized by $P(y_{i,k}/x_{i,k})$. Assuming that the PN sequences are generated randomly with equiprobable use of 1 and -1, it follows readily that the random variable $y_{i,k}$ has mean

$$E(y_{i,k}/x_{i,k} = m) = m\,(nP_i / \sum_{\substack{j=1 \\ j\neq i}}^{N} P_j)^{1/2}$$

and variance

$$E((y_{i,k} - \bar{y}_{i,k})^2/x_{i,k} = m) = 1$$

for m = 1 or -1. Since $y_{i,k}$ is the sum of a large number of random variables, the statistics of $y_{i,k}$ is approximately Gaussian by the law of large number. Hence by defining

$$N_{0,i}/2 = \sum_{\substack{j=1 \\ j\neq i}}^{N} P_j/n$$

and dropping the subscript k, we have the approximation

$$P(y_i/x_i = m) = \exp(-(y_i - m.\sqrt{2P_i/N_{0,i}})^2/2) / \sqrt{2\pi}$$

(Strictly speaking, the $y_{i,k}$'s are not mutually independent. This dependence, however, is weak and diminishes as the number of users N becomes large.)

The channel (actually the channel plus quantizer) can be modeled by the channel transition diagram in figure 6.1.1 and figure 6.1.2 for the cases of no quantization and hard quantization respectively. Figure 6.1.1 is essentially a binary

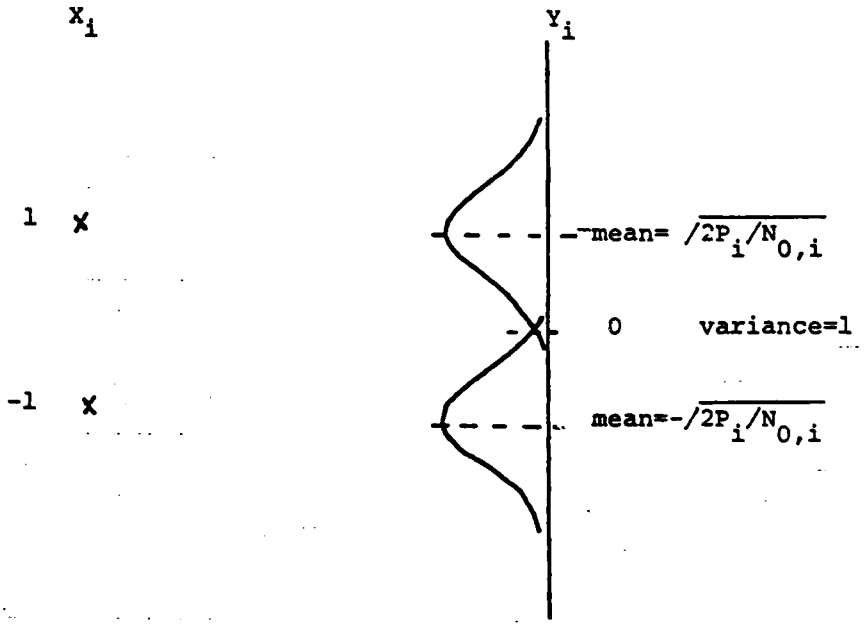


Figure 6.1.1 The binary input Gaussian output channel

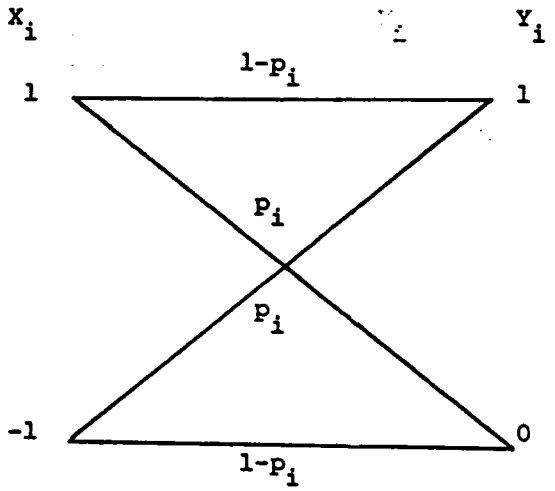


Figure 6.1.2 The binary symmetric channel

input Gaussian output channel, whereas figure 6.1.2 is a Binary Symmetric Channel (BSC) with cross-over probability $p=Q(\sqrt{2P_i/N_{0,i}})$ in which

$$Q(z) = \int_{-\infty}^{-z} e^{-x^2/2} dx/\sqrt{2\pi}$$

Consider the case of no coding on top of the PN sequence (i.e., $x_{i,k}=1$ if $u_{i,k} = 1$ and $x_{i,k} = -1$ if $u_{i,k} = 0$). The channel is therefore the BSC of figure 6.1.2. Assuming equal power for all users, the bit error probability upper bounded by

$$Q(\sqrt{2P_i/N_{0,i}}) = Q(\sqrt{n/(N-1)}) = Q(T^{-1/2})$$

in which T is the total throughput (N/n bits/chip). This upper bound is listed in figure 6.1.3. It is readily seen that the throughput has to be reduced substantially for tolerable error probability. An immediate conclusion is that coding should be used on top of PN sequence coding to achieve a lower bit error probability.

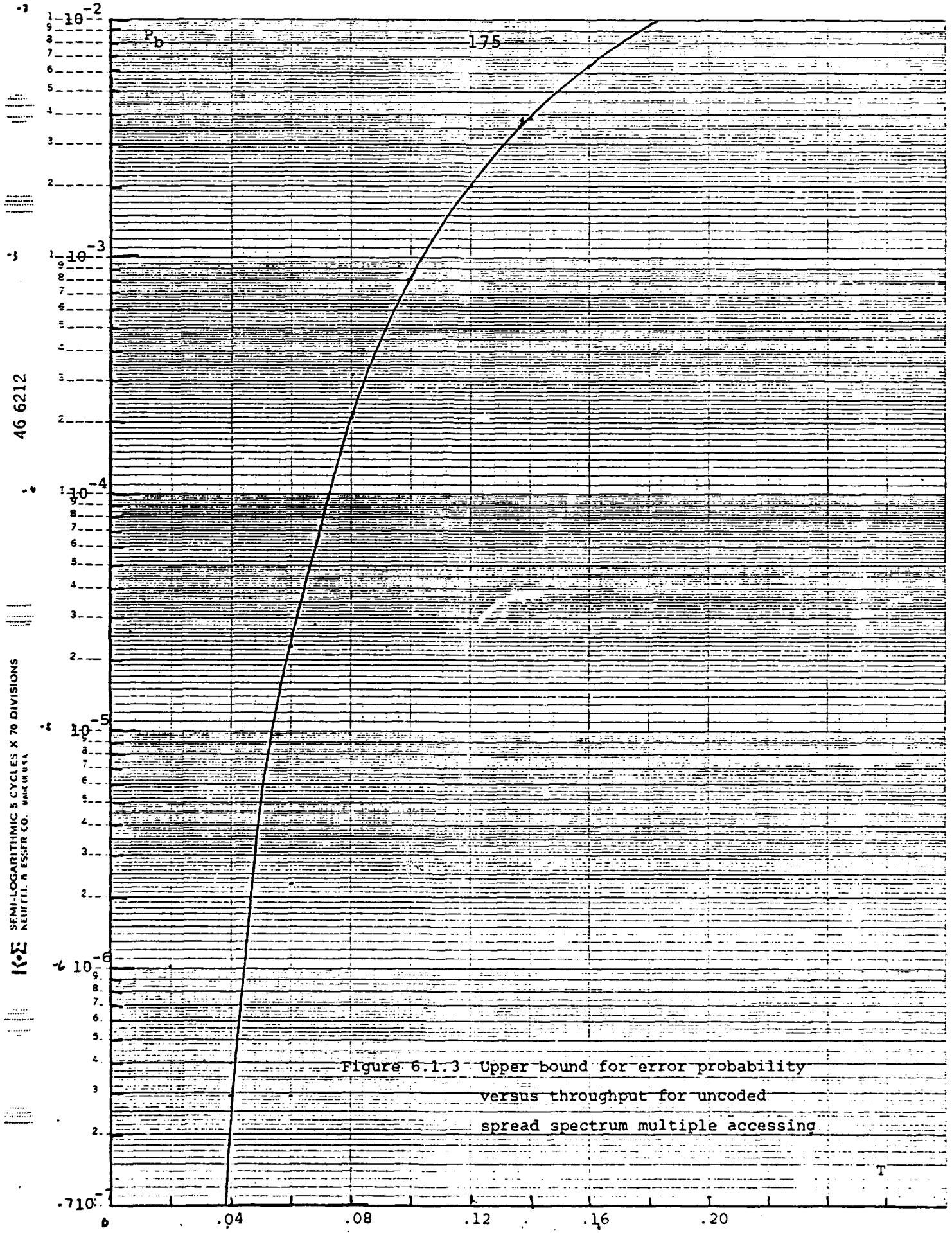


Figure 6.1.3 Upper bound for error probability versus throughput for uncoded spread spectrum multiple accessing

6.2 Capacity and cutoff rates

The capacity of a channel is the upper bound on the rate of reliable communication. The cutoff rate is the value of the random coding exponent function with $\rho = 1$ [9]. When sequential decoding is used, the cutoff rate is the maximum rate of communication over the channel with bounded average number of computation for decoding an information bit. For the BSC of figure 6.1.2, the cutoff rate and capacity (in bits/code symbol) are derived in [10], giving

$$R_{0,i} = 1 - \log_2(1 + \sqrt{4\, p_i(1 - p_i)})$$

$$C_i = 1 - H(p_i)$$

in which

$$p_i = Q(\sqrt{2P_i/N_{0,i}})$$

and H is the binary entropy function. For the binary input Gaussian output channel, we have [10]

$$R_{0,i} = 1 - \log_2(1 + e^{-P_i/N_{0,i}})$$

$$C_i = \log 2\pi e - \int_{-\infty}^{\infty} P^i(y) \log_2 P^i(y)\, dy$$

in which

$$P^i(y) = (P_{-1}^i(y) + P_1^i(y)) / 2$$

and

$$P_m^{i}(y) = P(Y_i = y_i / X_i = m) = \exp\left(-[y_i - m\sqrt{2P_i/N_{0,i}}\,]^2 / 2\right) / \sqrt{2\pi}$$

We shall assume henceforth that the transmitting power are equal for all users. Define $\beta = n/N$, the ratio of the length of the PN sequence to the total number of active user. Summing over the N users, and normalizing by n, the number of chips in a code symbol x_i, the sum-capacity and sum-cutoff rate (in bits per chip) for the BSC are

$$R_T(\beta) = N/n \quad R_{i,0} = \beta^{-1}\left(1 - \log_2(1 + \sqrt{4p(1-p)})\right)$$

$$C_T(\beta) = N/n \quad C_i = \beta^{-1}\left(1 - H(p)\right)$$

in which

$$p = Q\left(\sqrt{n/(N-1)}\right) \approx Q\left(\beta^{1/2}\right) \text{ for large N.}$$

For the binary input Gaussian channel,

$$R_T(\beta) = \beta^{-1}\left(1 - \log_2(1 + e^{-\beta/2})\right)$$

$$C_T(\beta) = \beta^{-1}\left(-\log_2 2\pi e - \int_{-\infty}^{\infty} P(y)\log_2 P(y)\,dy\right)$$

in which

$$P(y) = \left(P_1(y) + P_{-1}(y)\right) / 2$$

and

$$P_m(y) = \exp\left(-[y - m\beta^{1/2}] / 2\right) / \sqrt{2\pi}$$

These four functions of β are listed in figure 6.2.1 and 6.2.2. All four functions are observed to be monotonically decreasing in β. It can be readily shown that all four converges to the

Figure 6.2.1 R_T and C_T for the binary symmetric channel

β	$R_T(\beta)$	$C_T(\beta)$
	(bits/chip)	
0	.230	.459
0.1	.227	.448
0.2	.226	.441
0.3	.223	.429
0.4	.220	.418
0.5	.220	.413
0.6	.217	.403
0.8	.212	.384
1.0	.209	.369
1.2	.204	.353
1.4	.200	.338
1.6	.197	.326
1.8	.193	.314
2.0	.189	.301
2.5	.180	.274
3.0	.172	.250
4.0	.156	.211
6.0	.129	.157
8.0	.108	.122

Figure 6.2.2 R_T and C_T for the binary input Gaussian channel

β	$R_T(\beta)$ (bits/chip)	$C_T(\beta)$
0	.361	.721
0.05	.358	.704
0.1	.356	.688
0.2	.352	.657
0.4	.343	.604
0.6	.334	.559
0.8	.325	.520
1.0	.316	.486
1.2	.307	.454
1.4	.299	.428
1.6	.290	.403
1.8	.282	.381
2.0	.274	.361
2.5	.255	.317
3.0	.236	.282
4.0	.204	.288
5.0	.177	.190
6.0	.155	.162
8.0	.122	.124

function $1/\beta$ for large values of β. Asymptotic evaluation of these functions for vanishing values of β gives

$$R_T = \frac{1}{2\pi} \log e = .2296$$
$$C_T = \frac{1}{\pi} \log e = .4592$$

for the BSC and

$$R_T = \frac{1}{4} \log e = .3607$$
$$C_T = \frac{1}{2} \log e = .7213$$

for the binary input Gaussian channel.@ It is noteworthy that small values of β correspond to very noisy channels in figures 6.1.1 and 6.1.2, when the cutoff rate can be shown to be half the capacity. In fact, both R_T and C_T for diminishingly small β can be derived alternatively using a very noisy channel model [9,10]. The asymptotic result for large β suggests that using long PN sequences decreases the sum of the capacity and cutoff rate. The lesson is that we should use very low rate encoders and short PN sequences, so that β can be made small. The smallest value of β is achieved for $n=1$ which corresponds to using a very low rate code with no PN sequence coding.

@ It is well-known that hard quantization reduces the capacity by a factor of $\pi/2$ [9,10], as shown in this example.

6.3 Error probability and decoding complexity

This section studies the upper bound on error probability as a function of complexity. For constraint length k and rate 1/v convolutional codes, the ensemble average of the expected number of bit errors for an error event starting at a given time [10] is upper bounded by

$$P_b < 2^{-k v R_{o,i}} / [1 - 2^{-(v R_{o,i} - 1)}]^2$$

$$= 2^{-k v \beta R_T(\beta)} / [1 - 2^{-(v \beta R_T(\beta) - 1)}]^2$$

$$= [(1 + e^{-\frac{1}{2\beta v}})/2]^{v k} / [1 - 2((1 + e^{-\frac{1}{2\beta v}})/2)^v]^2$$

The last equality follows from

$$R_T(\beta) = \beta^{-1}(1 - \log_2(1+e^{-\beta/2}))$$

derived in section 6.2 for the binary input Gaussian channel, and the total throughput for the N users is

$$T = (N/n).(1/v) = 1/(v\beta)$$

This upper bound on P_b versus T, the degree of congestion of the channel, is shown in figure 6.3.1 for k=10 and v from 2 to 10. The result shows that using a large v gives far superior error probability for given k. Consequently, using a small β also improves the antijam capability of a user. By setting to zero the denominator of the upper bound on P_b, we have

$$T = -1/(2v \ln [2 (1/2)^{1/v} - 1])$$

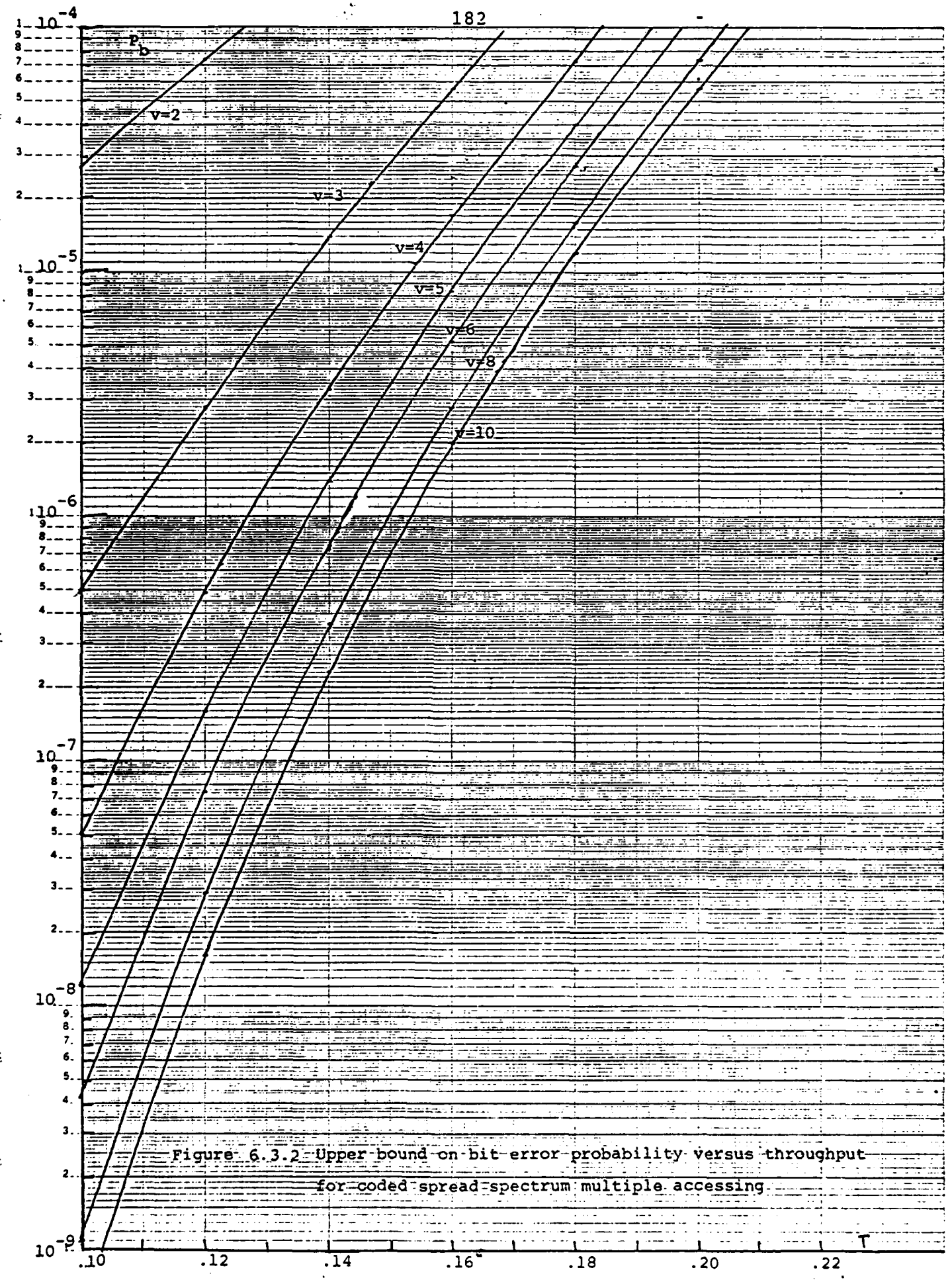


Figure 6.3.2 Upper bound on bit error probability versus throughput for coded spread-spectrum multiple accessing.

This value of T versus v is plotted in figure 6.3.2, together with the corresponding value of β (=1/(Tv)). It is observed that a large v (consequently a small β) increases the value of T. A value of v around 5 would give a value of T close to the maximum R_T.

Since $R_{0,\lambda}=\beta R_T(\beta)$, the error probability is decreasing at a rate that is exponential in $-nvkR_T(\beta)/N$. It is noteworthy that nv is the number of chips for the output on a transition of the trellis. Consider the use of Viterbi decoding, for which decoding complexity is measured by k (the trellis therefore has 2^{k-1} states), for fixed nv and N. Consider changing the value of β. (Since $\beta = n/N$ and nv is fixed, a smaller β would imply a smaller n and a larger v.) Due to the fact that $R_T(\beta)$ is maximized by small values of β, the complexity k is minimized for small values of β at a fixed error probability.

Using small β, however, may require a decoder complexity larger than that reflected by the value of k due to two reasons. First, decoding can be performed in units of code symbols since circuits on a chip are available for demodulating the entire PN sequence carrier (which is considered as a code symbol in the channel transition diagram). A small β would require a large v, the number of code symbols on each transition of the trellis. Second, the use of small β requires finer quantization for decoding since the channel (figure 6.1.1) becomes more noisy. The merging of the four functions in figures 6.2.1 and 6.2.2 in fact shows that rough quantization may be used with little effect on R_T

Figure 6.3.2 Throughput T for given v

v	β	T (bits/chip)
2	1.765	.284
3	1.066	.313
4	0.766	.326
5	0.599	.334
6	0.493	.339
7	0.418	.342
8	0.363	.344

or C_T for large β .

Since user interference is severe in such system, using a long constraint length code is crucial for low error probability. Unfortunately, Viterbi decoding is too complex to implement for constraint lengths beyond 10. We suggests searching for good convolutional codes to decrease error probability and increase total throughput. (Our previous bound on P_b is a random coding bound for an ensemble of codes chosen at random.) A directory of good encoders would be built up, from which each subscriber would choose an encoder. The maximum size of the directory, which determines the number of subscribers for the system, depends on the density of good codes times the volume of the code space. We believe that searching for good convolutional encoders [17] is a better and more tractable way to improve code orthogonality than trying to design PN sequences with good auto-correlation and low cross-correlation, which is difficult due to code asynchronism.

A long PN sequence may also make synchronization at the receiver easier to achieve. It should be noted that two synchronizations have to be performed, namely the synchronization of the PN sequence carrier and the synchronization of the convolutional encoder, up the the code symbols. For the case of n=1, there is no need for PN sequence synchronization. In practice, a synchronization signal can be sent on top of the coded signal for easy code synchronization.

Priorities among the users can be set up by the authorization of transmitting power levels for the users. Using a

higher transmitting power, a user may choose to increase its data rate or decrease its error rate. The analysis in this chapter can be carried further to show that the capacity for each user is directly proportional to its transmitting power. Fading, due to distance and poor weather condition, can also cause variations of signal power and results in dramatically deteriorated system performance. The system, in order to operate reliably in most conditions, would have to include a power margin that reduces the total throughput of the system. Also, power monitoring for fairness is difficult in practice.

It has been suggested [5] that CDMA for the spread spectrum channel should operate at a throughput of about 10 percent for reasonable error performance. The analysis in this paper gives a maximum throughputof 36 percent based on the cutoff rate. In practice, the throughput would have to be reduced for a smaller decoding complexity required for a tolerable bit error rate. We have shown that a throughput of 20 percent is feasible in practice.

Chapter 7 Multiple Accessing for the Collision Channel

For the collision channel, each user sends packets, which are destroyed if they overlap. We shall consider the use of such a channel for multiple accessing, when each user redundantly encodes the information to be sent. The resulting data sequence is blocked into packets for transmission. The receiver collects all the packets from the corresponding transmitter and ignores all other packets and collisions. These collected packets are subsequently decoded. Section 7.1 describes the channel model and derives the capacity region. Section 7.2 analyses the performance of block codes. Section 7.3 proposes a convolutional coding and interleaving scheme, and compute the sum-cutoff rate of the channel. Section 7.4 analyses a decoding scheme that keep track of all active paths in the trellis and evaluates its decoding complexity. Section 7.5 generalizes sections 7.1 and 7.3 to the case when the channel is corrupted by additive Gaussian noise. It shows that substantial power saving is gained, at hardly any extra cost, by the redundant coding that is originally intended to correct erasures due to packet collisions.

7.1 Modeling and channel capacity

The slotted (or packet synchronized) noiseless collision channel is defined as follows. The code symbol alphabet is $X_i = \{0,1,2,\ldots,2^n\}$ for each user, in which 0 represents an idle and the letters 1 to 2^n each represents a packet of length n bits. The channel transition is given by $y=x_j$ if $x_i=0$ for all $i\neq j$, and y=erasure if more than one $x_i\neq 0$. The channel transition diagram for user i is shown in figure 7.1.1.

Without code synchronization nor joint decoding, the capacity region $\mathcal{R}$ is given by

$$\mathcal{R} = \bigcup_{P(x_1\ldots x_M y)} \tilde{\mathcal{R}}$$

$$:P(x_1\ldots x_M y)=P_1(x_1)\ldots P_M(x_M)P(y/x_1\ldots x_M)$$

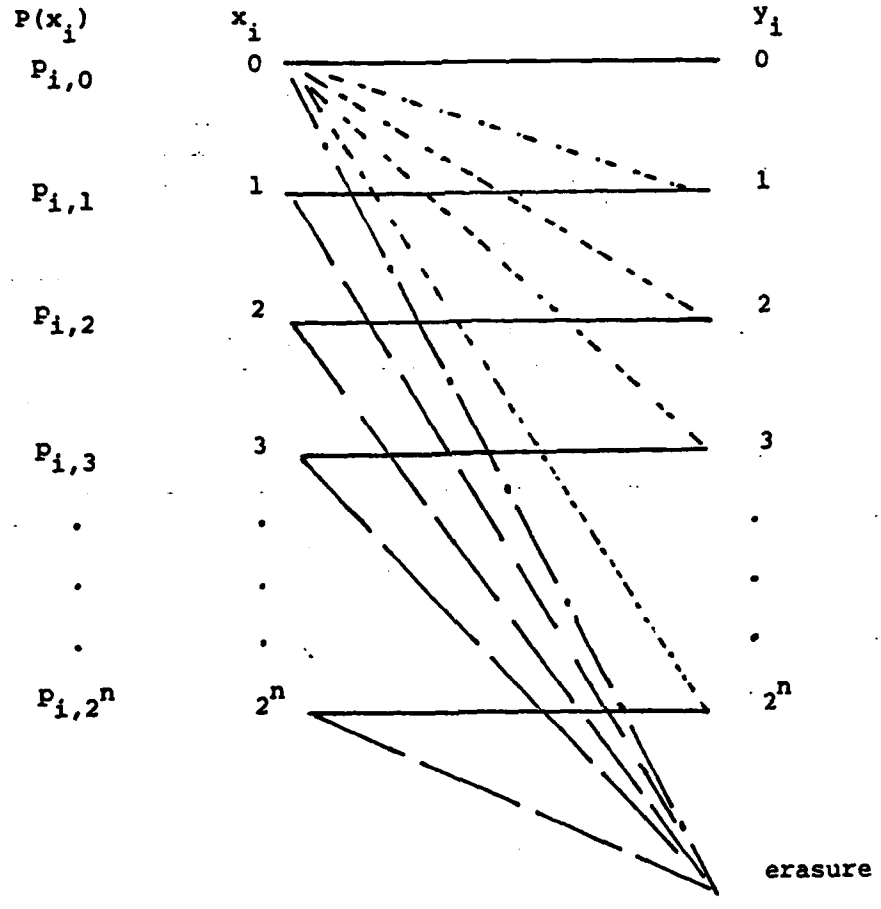
in which $(R_1,\ldots,R_M)\in\tilde{\mathcal{R}}$ iff

$$R < I(X_i;Y) \text{ for all } i.$$

Due to the symmetry of the letters 1 to 2^n, it is apparent that each of the 2^n nonzero packets should be equiprobable for achieving capacity. Therefore, we shall abbreviate $p_{i,0}$ by p_i, and $p_{i,j}$ by $(1-p_i)/2^n$ for all $1\leq j\leq 2^n$. It can be shown after some tedious evaluation that

$$I(X_i;Y) = n\,(1-p_i)/p_i \prod_{j=1}^{M} p_j + 0(1) \quad \text{bits/slot}$$

$$\cong (1-p_i)/p_i \prod_{j=1}^{M} p_j \quad \text{n-bit/slot}$$



Transition probabilities

——————— P(no packet sent by the other M-1 users)

- - - - - - - - - P(exactly one other user sends the message y_i)

——— · ——— P(two or more packets sent by the other M-1 users)

———— ——— — P(one or more packets sent by the other M-1 users)

Figure 7.1.1 Channel transition diagram for the collision channel

for large n. We would like to have a characterization of $\mathcal{R}$ without the p_i's. Consider the equations

$$R_i = (1-p_i)/p_i \prod_{j=1}^{M} p_j$$

for all i. Differentiating the R_i's with respect to the p_i's gives

$$dR_i = \sum_{\substack{k=1 \\ k\neq i}}^{M} \{ (1-p_i)/(p_i p_k) \prod_{j=1}^{M} p_j \} dp_k - (\prod_{j=1}^{M} p_j)/p_i \; dp_i$$

which in matrix form can be expressed as

$$\underline{dR} = \begin{bmatrix} dR_1 \\ dR_2 \\ . \\ . \\ dR_M \end{bmatrix} = \begin{bmatrix} 1/p_1 & 0 & \cdots & 0 \\ 0 & 1/p_2 & \cdots & 0 \\ . & . & \cdots & . \\ . & . & \cdots & . \\ 0 & 0 & \cdots & 1/p_M \end{bmatrix} \cdot \begin{bmatrix} -p_1 & 1-p_1 & 1-p_1 & \cdots & 1-p_1 \\ 1-p_2 & -p_2 & 1-p_2 & \cdots & 1-p_2 \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 1-p_M & 1-p_M & 1-p_M & \cdots & -p_M \end{bmatrix} \cdot$$

$$\begin{bmatrix} 1/p_1 & 0 & \cdots & 0 \\ 0 & 1/p_2 & \cdots & 0 \\ . & . & \cdots & . \\ . & . & \cdots & . \\ 0 & 0 & \cdots & 1/p_M \end{bmatrix} \cdot \begin{bmatrix} dp_1 \\ dp_2 \\ . \\ . \\ dp_M \end{bmatrix}$$

At the boundary of $\mathcal{R}$, it is necessary that the vector $(dR_1, dR_2, \ldots, dR_M)$ can only be tangential to the surface of the boundary for all vectors $(dp_1, dp_2, \ldots, dp_M)$. Therefore, it is necessary that the second matrix on the right hand side be singular. Let $\underline{c}_i$ be the i-th column vector of that matrix. Linear dependence of these vectors implies that $\sum_i \alpha_i \underline{c}_i = \underline{0}$ for some α_i's which are not all zeros. Consequently,

$$\begin{bmatrix} (\sum_i \alpha_i)(1-p_1) - \alpha_1 \\ (\sum_i \alpha_i)(1-p_2) - \alpha_2 \\ \vdots \\ (\sum_i \alpha_i)(1-p_M) - \alpha_M \end{bmatrix} = \underline{0} \qquad *$$

Summing the elements of the above vector, we have

$$(\sum_i \alpha_i)(\sum_j (1-p_j)) - (\sum_j \alpha_j) = 0$$

which implies

$$q_1 + q_2 + \ldots + q_M = 1 \qquad (\text{as } \sum_i \alpha_i \neq 0 \text{ from } *)$$

where $q_i = 1-p_i$. Thus, the sum of the probabilities of sending a packets for the M users equals one at the boundary of $\mathcal{R}$. Therefore, the boundary of $\mathcal{R}$ satisfies the following set of equations,

$$\sum_{i=1}^{M} (1-p_i) = 1$$

$$R_i = (1-p_i)/p_i \prod_{j=1}^{M} p_j \quad , \quad 1 \leq i \leq M$$

For M=2, the elimination of p_1 and p_2 from the above equations gives

$$\sqrt{R_1} + \sqrt{R_2} = 1$$

which is shown in figure 1.2.1 d. It is very difficult to solve the above equations in useful forms for M>2. Instead, we shall obtain the sum of the rates along the main diagonal (by having

equal p_i 's). The sum has a maximum value of $(1-1/M)^{M-1}$. For large M, this value converges to e^{-1}, which is the same as the capacity of the slotted Aloha channel.

The capacity region for the case with joint decoding (given in chapter 2) is the same as that without joint decoding if the length of packet n is large enough. This is due to the fact that collisions result in erasures that are totally useless for decoding, as well as the fact that the packets for the other messages may be ignored without much loss in the decoding of one's message.

The remainder of this section examines how unslotted packets affect the capacity of the channel. Let $1-p_i$ be the probability that user i transmits a packet in the slot. The probability (conditioned on a transmission in a slot) that a packet for user i does not collide with the packets of user j is p_j^2, because the packet for user i may collide with a packet in the two slots that overlap with the packet of user i. Hence the throughput for user i is given by

$$T_i = (1-p_i)/p_i^2 \prod_{j=1}^{M} p_j^2$$

Assume equal rates for all users so that $p_i=p_j=p$; the sum of the rates is then

$$T = \sum_{i=1}^{M} T_i = M\,(1-p)\,p^{2(M-1)}$$

which is maximized when $p=1-1/(2M-1)$, with

$$T = M/(2M-1)\ \ (1-1/(2M-1))^{2(M-1)}$$

For large M, T approaches $e^{-1}/2$, which is the same as the capacity for the pure Aloha scheme.

In practice, we may be able to recover the front part of a packet up to the point when it starts to collide with another packet, as shown in figure 7.1.2. No portion of a packet which starts at a time when another packet is being transmitted may be recovered, since the preamble (for identification and receiver synchronization), which is placed at the beginning of the packet, is lost in the collision. We assume the length of the preamble to be small compared with the length of the packet. Appendix 7.1 shows that the maximum sum of the throughput equals 1/4.

Massey [6] has shown that the capacity of the unslotted channel is the same as that of the slotted channel. Consider the grouping of u packets into a super-packet as shown in figure 7.1.3. The key idea is that even though a packet is totally lost through partial overlapping with other packets, part of a super-packet can be retrieved when it overlaps partially with other super-packets, as shown in figure 7.1.3. The proof of the e^{-1} sum-throughput is shown in Appendix 7.2.

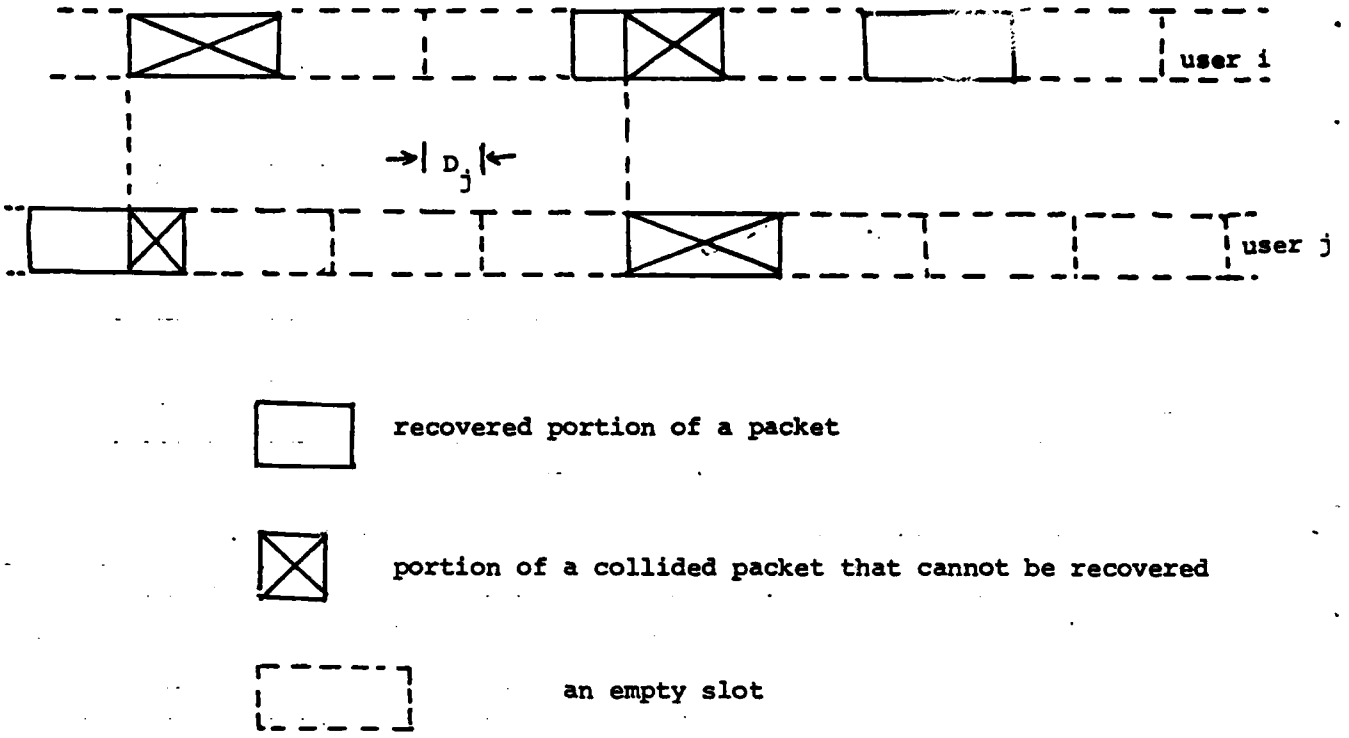


Figure 7.1.2 Partially recoverable packets

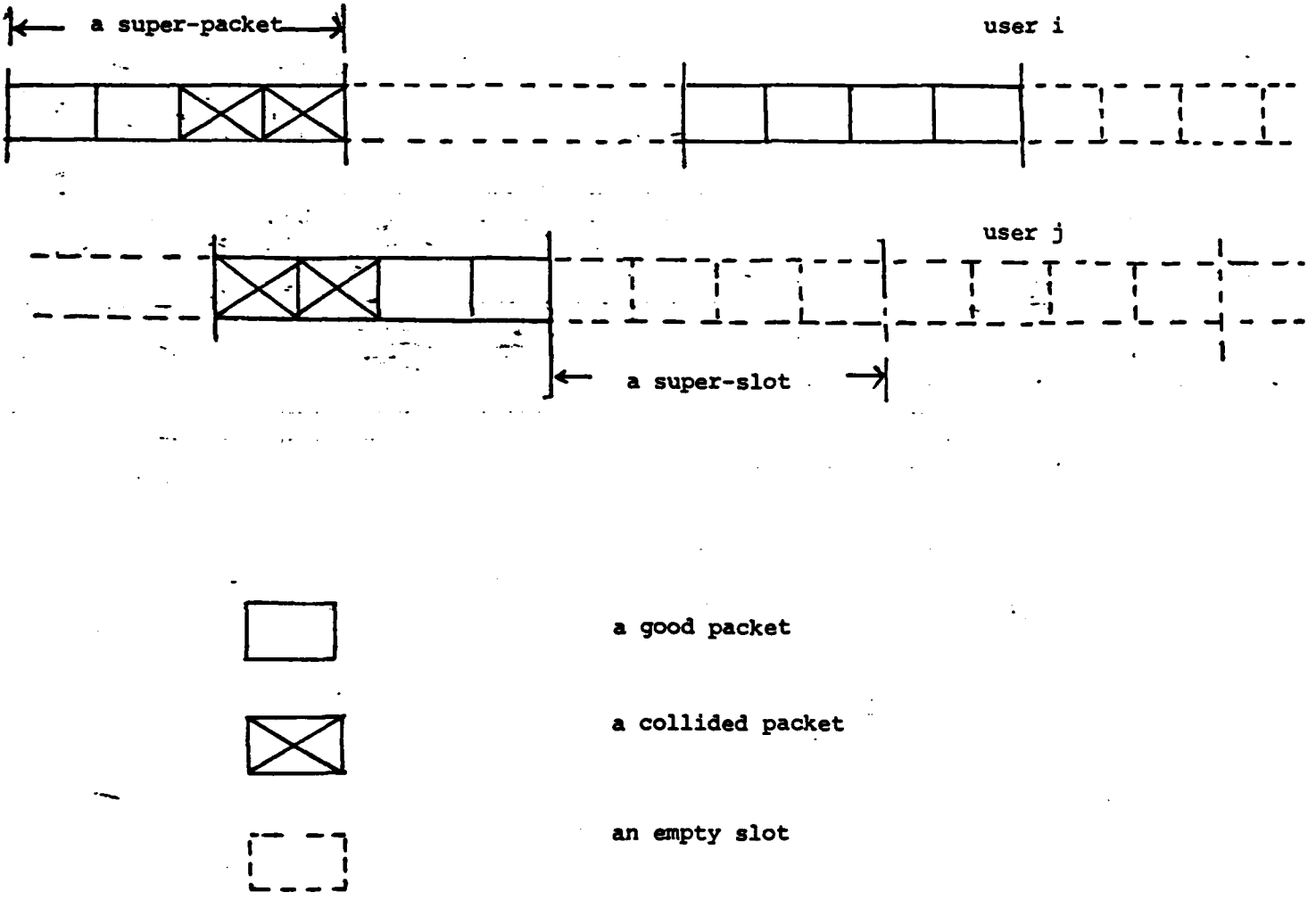


Figure 7.1.3 Super-packets for achieving the e^{-1} capacity

7.2 Block coding scheme

This section examines the use of block codes for the slotted collision channel. At the end of the section, we show how super-packeting may be used for the unslotted channel.

Reed-Solomon codes are especially effective for the correction of erasures. Each packet, which consists of n bits, may be treated as an element of the Galois field $GF(2^n)$. A Reed-Solomon code, defined on $GF(2^n)$, has codewords $f=(f_1,\ldots,f_k)$ satisfying the equations (for some integers m, d, and distinct $z_1,z_2,\ldots,z_k \in GF(2^n)$.)

$$\underbrace{\begin{bmatrix} z_1^m & z_2^m & \cdots & z_k^m \\ z_1^{m+1} & z_2^{m+1} & \cdots & z_k^{m+1} \\ z_1^{m+2} & z_2^{m+2} & \cdots & z_k^{m+2} \\ \vdots & \vdots & \cdots & \vdots \\ z_1^{m+d-2} & z_2^{m+d-2} & \cdots & z_k^{m+d-2} \end{bmatrix}}_{Z} \underbrace{\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_k \end{bmatrix}}_{f} = \underline{0}_{d-1}$$

Consequently, the Reed-Solomon code has code rate $(n-d+1)/n$. Define the distance between two codewords $f=(f_1,\ldots,f_k)$ and $f'=(f_1',\ldots,f_k')$ as the total number of places where $f_i \neq f_i'$. We shall give a proof for the well-known fact that this Reed-Solomon code has distance at least d.

Proof

Suppose this is not true. Then there exist two codewords f

and f' such that f''= f + (-f') has less than d nonzero entries. Since $Zf''=\underline{0}$, it follows that there exists a linear combination of less than d of the column vectors of Z which equals the zero vector. This is impossible because Z is of full rank, namely d-1, which follows from the fact that Z is a Vandermonde matrix with

$$|Z| = \left[\prod_{i=1}^{k} z_i^m\right] \left[\prod_{\substack{i>j \\ i,j\in\{1,\cdots,k\}}} (z_i - z_j)\right] \neq 0$$

since the z_i's are distinct.

Q. E. D.

Massey [6] used the Reed-Solomon code to prove that zero error probability can be achieved with each user transmitting at an information rate of $(1-1/M)^{M-1}/M$ n-bit per slot. The scheme involves a clever way for each user to put M^{M-1} packets in a time frame of M^M slots, such that exactly $(M-1)^{M-1}$ packets would be collision free for each user, no matter how the frames of the users are relatively shifted (slot-wise). Therefore, a Reed-Solomon code of rate $(M-1)^{M-1}/M^{M-1}$ would suffice to correct all the erasures. The throughput for each user is then $(M-1)^{M-1}/M^M = (1-1/M)^{M-1}/M$. The sum-throughput is $(1-1/M)^{M-1}$, which is the same as the capacity derived in section 7.1.

The scheme of Massey is inadequate for implementation if the number of users (M) is large or variable. However, Reed-Solomon codes are still effective for such a channel. Assume that a user puts a packet in a slot with probability q=1-p and the k consecutive packets he puts on the channel comprise a

codeword for a rate r=h/k Reed-Solomon code. Consequently, the information rate for the user is q.h/k n-bit per slot. The sum-throughput is T=Mqr n-bit per slot. The probability of erasure for a packet is

$$\epsilon = 1 - (1-q)^{M-1} = 1 - (1 - T/Mr)^{M-1} \approx 1 - e^{-T/r}$$

for large M. The probability of block error is upper bounded by

$$P_b \leq \sum_{i=h}^{k} {}_kC_i \; \epsilon^i (1-\epsilon)^{k-i}$$

Using an upper bound [9] on the sum of the tail of a binormial distribution , we have

$$\begin{aligned} & P_b \\ \leq\ & h\,(1-\epsilon)/(h(1-\epsilon) - (k-h)\epsilon\,) \;\; {}_kC_h \; \epsilon^h (1-\epsilon)^{k-h} \\ \leq\ & r\,(1-\epsilon)/(r(1-\epsilon) - (1-r)\,\epsilon\,) \;\; e^{k H_\epsilon(r)} \; \epsilon^h (1-\epsilon)^{k-h} \\ =\ & r\,(1-e^{-T/r})/(re^{-T/r} - (1-r)(1-e^{-T/r})\,) \;\; [r^{-r}(1-r)^{-(1-r)}(1-e^{-\frac{T}{r}})^r \; e^{-\frac{T}{r}(1-r)}]^k \end{aligned}$$

This upper bound $P_e(T,r,k)$ is plotted as a function of T for several codes in figure 7.2.1. The error probability is not very satisfactory for these codes. We shall show that convolutional codes have a more superior error probability.

If the channel is unslotted, groups of u Reed-Solomon code symbols may be grouped together to form a superpacket. For large u, the probability of erasure for a particular packet is the same as that of the unslotted channel. Thus the throughput of the

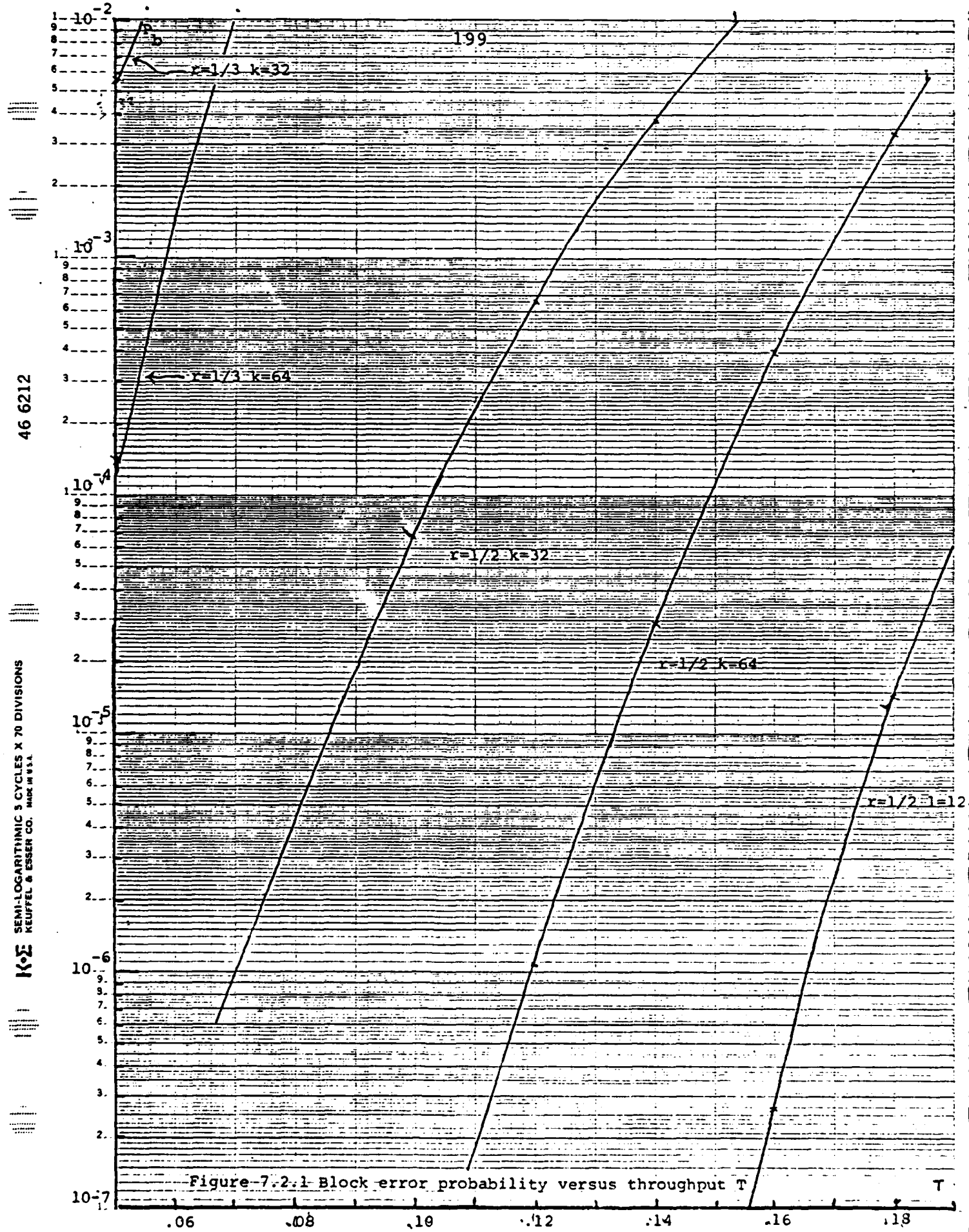


Figure 7.2.1 Block error probability versus throughput T

unslotted channel is the same as the e^{-1} throughput of the slotted channel.

7.3 Convolutional coding scheme

We now turn to the use of convolutional codes. The encoding scheme is shown in figure 7.3.1. The outputs of the rate r=1/v constraint length k code (v=2 in the figure) is fetched alternately by the switch S_1 . Switch S_2 puts successive bits into successive packets in a stack of m packets. When all m packets are filled, each packet would be transmitted after a random delay.

For the unslotted channel, the data sequence is fed in parallel (figure 7.3.2) into u encoder-interleavers (the square box in figure 7.3.1), and the u packets from the encoder-interleavers are grouped together to form a super-packet. The super-packet is transmitted after a random delay.

Decoding delay results from the fact that the data bits have to be deinterleaved before decoding. This delay is of the order mn. To minimize delay, we may use a smaller stack or shorten the length of a packet. Shortening the length of a packet, however, would incur a higher fraction of preamble overhead in the message body. The number of packets in the stack should be at least k.v, so that each bit in a packet is generated from a different set of information bits. If m<k.v, error probability would increase because erasures are clustered together. Choosing m slightly larger than k.v would suffice since the likelihood of long error paths in the trellis is small.

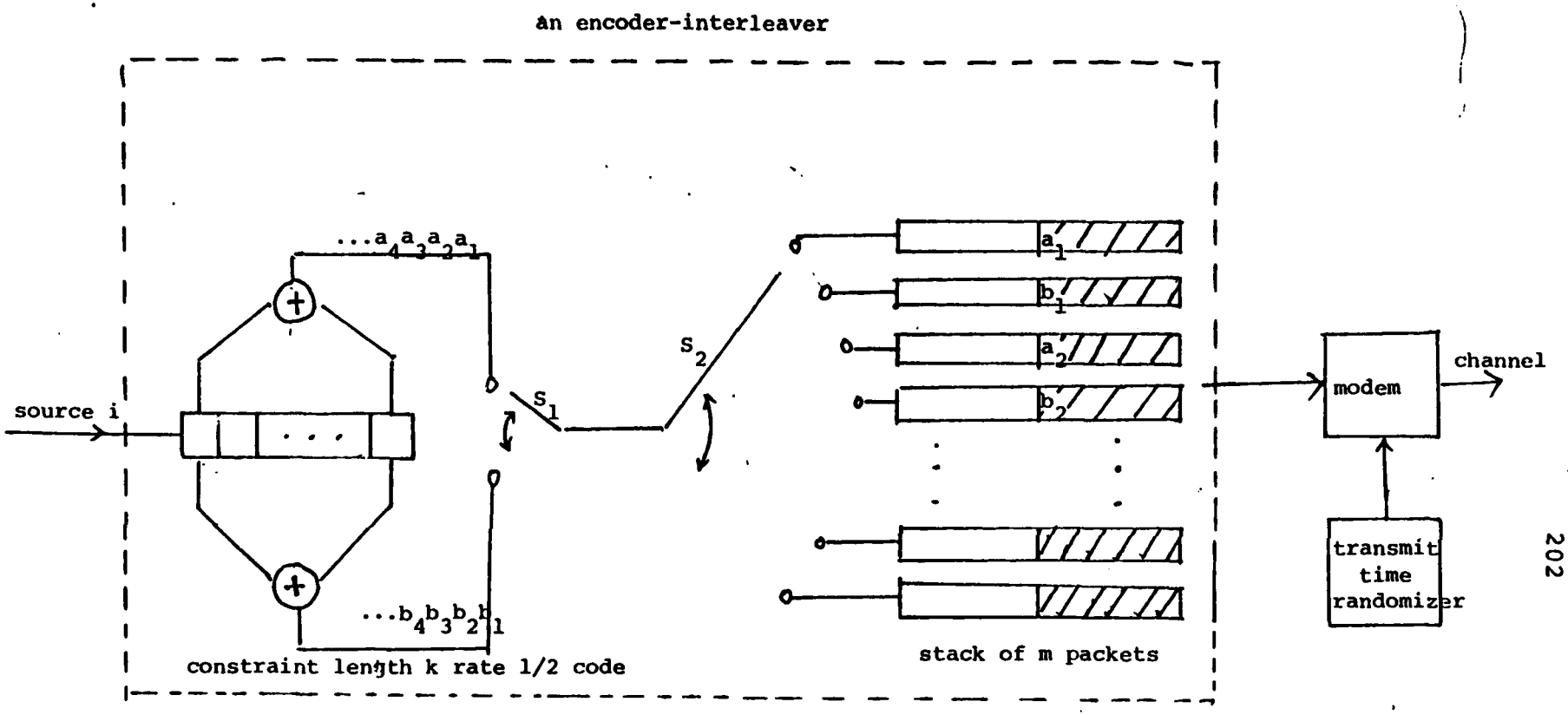


Figure 7.3.1 Multiple access schemes for convolutional codes

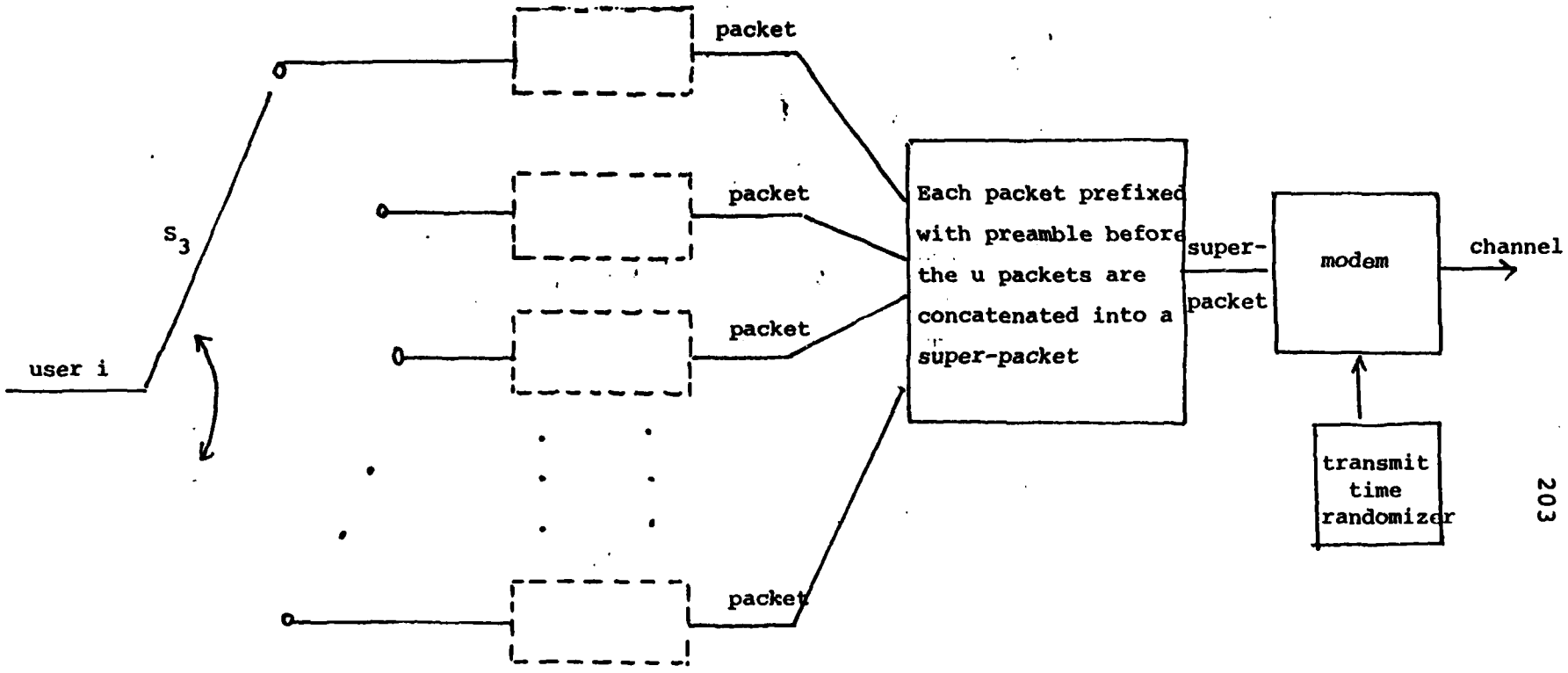


Figure 7.3.2 Formation of super-packets

Decoding delay due to deinterleaving can be reduced by the use of convolutional interleaving as shown in figure 7.3.3. Each packet is filled with n/m more bits than the next packet in the stack. The packet at the top of the stack is popped off the stack when the packet is filled. An empty packet is added to the bottom of the stack when the packet at the top is popped. The popped packet is transmitted after a random delay, which is smaller than the packet interarrival time. Using this scheme, delay incurred and buffer required are about half that of the previous interleaving scheme. The convolutional interleaving scheme, however, has a problem with the first few packets of a message, since the packet at the top of the stack is empty at the beginning. It seems that the first few packets would have to be filled in the manner shown in figure 7.3.4, subsequently wasting m/2 packets per message. This waste is tolerable only if the message is long. The choice between these two schemes depends on the length of the message and the amount of decoding delay that can be tolerated.

We now give a random coding bound [10], based on the cutoff rate, for the error probability. Consider the deinterleaved data sequence at the receiver. The occurence of erasures in the deinterleaved sequence is not memoryless in the sense that erasures, if they occur, are periodic. If we use a decoding metric that is time- invariant and a convolutional encoder of long constraint length, these erasures may be viewed as being memoryless. Hence, we may treat the channel as a binary erasure channel shown in figure 7.3.5 if we impose the

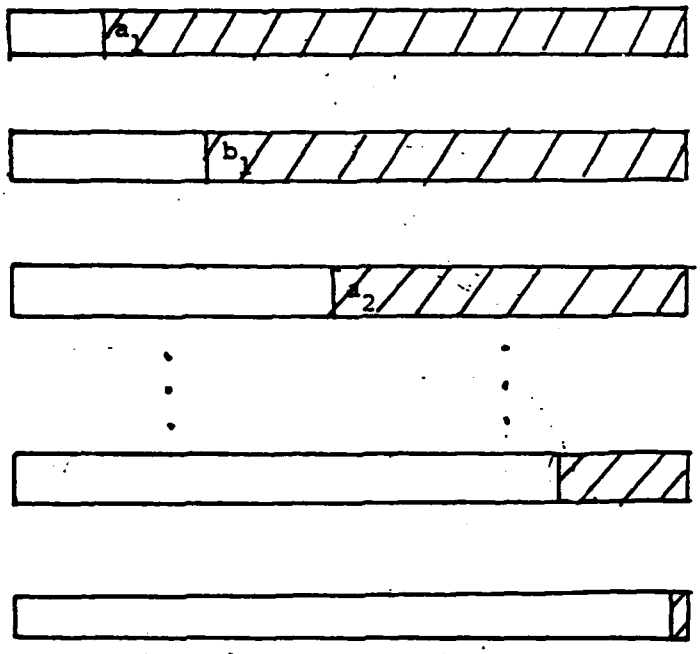


Figure 7.3.3 Convolutional interleaving

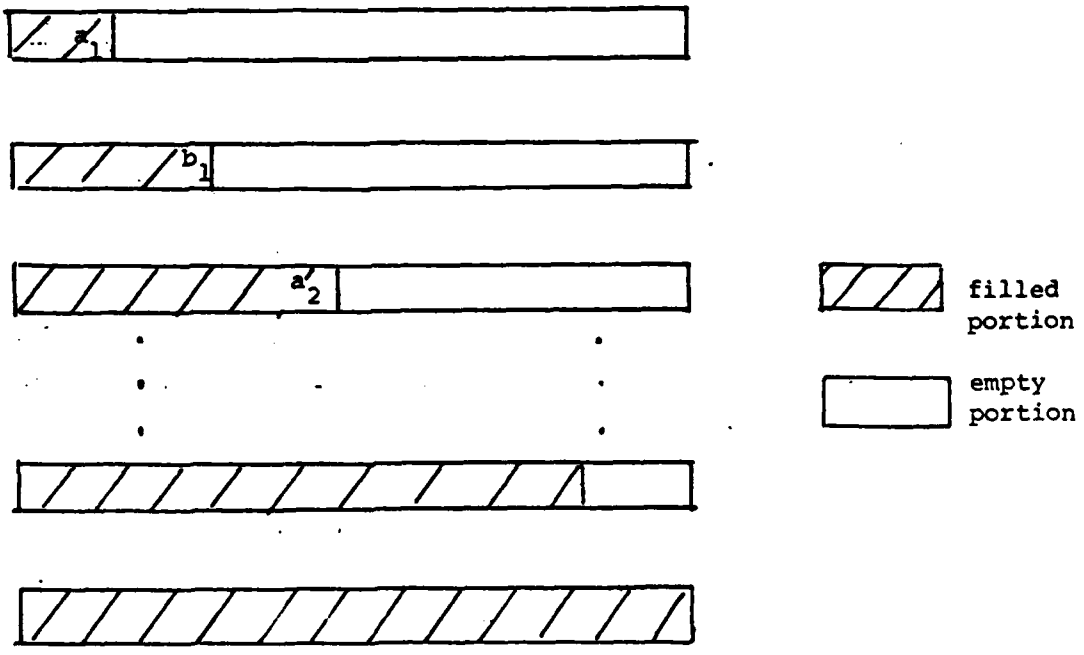


Figure 7.3.4 Length of the first few packets of a message

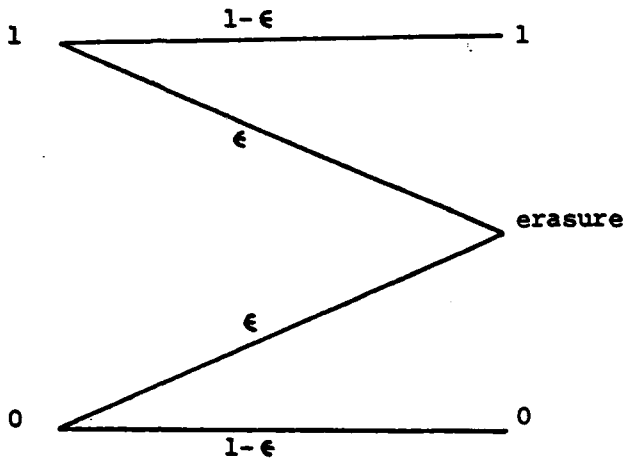


Figure 7.3.5 The binary erasure channel

time-invariant structure on the decoder. The capacity of the binary erasure channel is equal to $C=1-\epsilon$. Consequently, the sum-capacity is

$$C_T = \max_{q} \; M q (1-\epsilon) = \max_{q} \; M q (1-q)^{M-1} = (1-1/M)^{M-1}$$

which approaches e^{-1} for large M. The cutoff rate for the binary erasure channel is

$$R_0$$

$$= -\log_2 \left(\sum_y \left(\sum_x P(x) \sqrt{P(y/x)} \right)^2 \right)$$

$$= -\log_2 \left((1 + \epsilon)/2 \right)$$

$$= -\log_2 (1 - e^{-Tv}/2)$$

since $\epsilon = 1-e^{-Tv}$ for large M. The sum of the cutoff rates for the M users is

$$R_T = M q R_0 = - Tv \log_2 (1 - e^{-Tv}/2)$$

since $Mq/v=T$. Reliable communication using sequential decoding is achievable for T up to R_T. Consequently the maximum sum-throughput satisfies

$$T = - Tv \log_2 (1 - e^{-Tv}/2)$$

or

$$T = -1/v \; \ln (2 - 2 \times 2^{-\frac{1}{v}})$$

T has a maximum value of 0.295096 when v=3.014. For practical purposes, v is an integer so the best code rate is 1/3, with

T=0.295093. For v=2, the maximum T is 0.2674.

Errors in decoding can arise when an incorrect path with a higher metric merges into the correct path. The ensemble average of the expected number of bit errors for such an error event starting at a given time is upper bounded [10] by

$$P < 2^{-vR_0k} / [1-2^{-(vR_0-1)}]^2$$

$$= (1-e^{-Tv}/2)^{vk} / [1-2(1-e^{-Tv}/2)^{v}]^2$$

This bound is plotted in figure 7.3.6 for several code rates and constraint length. The plot shows that long constraint length should be used to combat the severe user interference. Such long constraint lengths make Viterbi decoding impractical. Thus the decoder should keep track of only a subset of states, the size of which is independent of the constraint length. The error probability for such decoders is determined by the size of the buffer of the decoder. The next section investigates the probability of buffer overflow versus buffer size of the decoder for a particular decoding scheme.

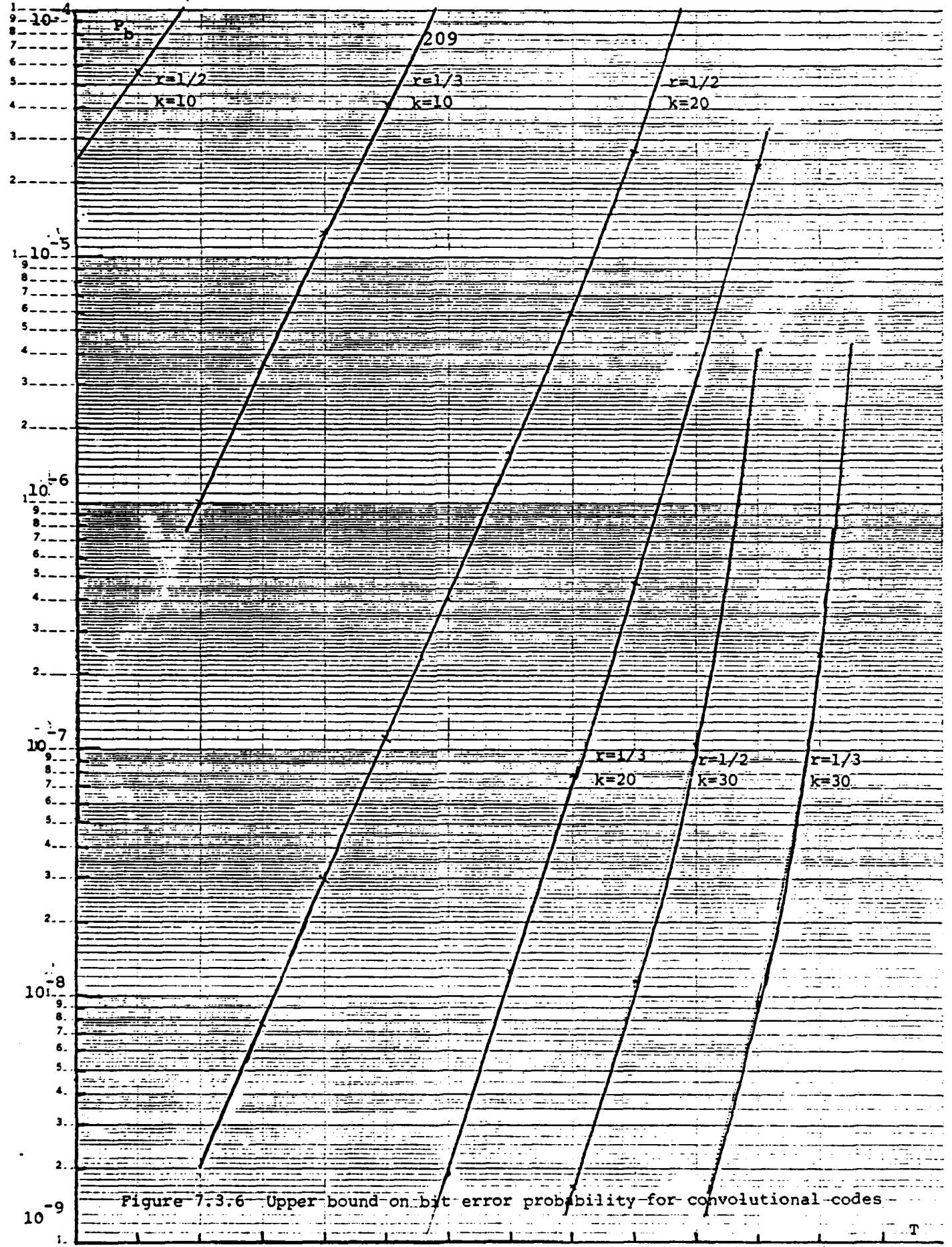


Figure 7.3.6 Upper bound on bit error probability for convolutional codes

7.4 Decoding complexity

This section examines a decoding scheme for the binary erasure channel and demonstrates its small decoding complexity at a sum-throughput that is close to the sum-cutoff rate of $-1/v \cdot \ln(2-2X2^{-\frac{1}{v}})$.

The decoder keeps a list of the active states of the trellis. A state is active if there is a path leading up to the state such that the code sequence of the path is agreeable with the deinterleaved channel sequence. The code symbol zero (or one) is disagreeable with the channel symbol one (or zero) while the channel symbol erasure is agreeable with the code symbols zero or one. The complexity of decoding is measured by the random variable n, the number of active states in steady state, with probability distribution P(n). The number of active states is upper bounded by treating the trellis as a tree, with the merged states double counted. Since long constraint length codes are used, the trellis in fact resembles a tree for short paths off the correct path. Long incorrect paths are highly unlikely, thus decoding error due to the merging of an incorrect path with the correct path may be ignored.

Let n_j be the number of active states at time j of the trellis. Let $e_{t,j}$ be the number of states at time j+1 that are made active from the active state t at time j. We shall drop the superfluous subscript j for $e_{t,j}$. Without loss of generality, the correct path is assumed to traverse the state sequence of all

ones, and the plausible states are labeled from 2 to n_j. Thus

$$n_{j+1} = \sum_{t=1}^{n_j} e_t$$

with probability generating function

$$\begin{aligned}
& S_{j+1}(s) \\
&= \sum_{k=0}^{\infty} P(n_{j+1}=k)\, s^k \\
&= \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} P(n_{j+1}=k/n_j=m)\, P(n_j=m)\, s^k \\
&= \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} P(n_j=m)\, P(e_1+\ldots+e_m=k)\, s^k \\
&= \sum_{m=0}^{\infty} P(n_j=m) \sum_{k=0}^{\infty} P(e_1+\ldots+e_m=k)\, s^k
\end{aligned}$$

in which the probability generating function

$$\begin{aligned}
& \sum_{k=0}^{\infty} P(e_1+\ldots+e_m=k)\, s^k \\
&= \sum_{r=0}^{v} P(r \text{ erasures in } v \text{ channel symbols}) \times \\
&\qquad \sum_{k=0}^{\infty} P(e_1+\ldots+e_m=k \;/\; r \text{ erasures in } v \text{ channel symbols})\, s^k \\
&= \sum_{r=0}^{v} {}_vC_r\, \epsilon^r (1-\epsilon)^{v-r} \prod_{t=1}^{m} V_{r,t}(s)
\end{aligned}$$

where

$$V_{r,t}(s) = \sum_{i=0}^{\infty} P(e_t = i/r \text{ erasures})\, s^i$$

is the probability generating function of e_t conditioned on the number of erasures. The product $\prod_{t=1}^{m} V_{r,t}(s)$ results from the fact that the random variables e_t given r erasures are independent for various t's and that the probability generating function of a sum of independent random variables is equal to the product of the

generating function of each random variable, which is given by

$$V_{r,1}(s)$$

$$= P(e_1=1/r \text{ erasures})\, s + P(e_1=2/r \text{ erasures})\, s^2$$

$$= (1-2^{-(\nu-r)})\, s + 2^{-(\nu-r)} s^2$$

and

$$V_{r,t}(s)$$

$$= P(e_t=0/r \text{ erasures}) + P(e_t=1/r \text{ erasures})\, s +$$

$$P(e_t=2/r \text{ erasures})\, s^2$$

$$= (1-2^{-(\nu-r)})^2 + 2 \times 2^{-(\nu-r)}(1-2^{-(\nu-r)})\, s + 2^{-2(\nu-r)} s^2$$

$$= (1 - 2^{-(\nu-r)} + 2^{-(\nu-r)} s)^2$$

for $2 \le t \le m$. Consequently,

$$S_{j+1}(s)$$

$$= \sum_{m=1}^{\infty} P(n_j=m) \sum_{r=0}^{\nu} {}_{\nu}C_r\, \epsilon^r (1-\epsilon)^{\nu-r}\, s\, (1 - 2^{-(\nu-r)} + 2^{-(\nu-r)} s)^{2m-1}$$

$$= \sum_{r=0}^{\nu} {}_{\nu}C_r\, \epsilon^r (1-\epsilon)^{\nu-r}\, s/(1 - 2^{-(\nu-r)} + 2^{-(\nu-r)} s)$$

$$S_j\left((1 - 2^{-(\nu-r)} + 2^{-(\nu-r)} s)^2 \right)$$

The steady state probability generating function $S(s)$ satisfies the above equation after dropping the subscripts j and j+1. Differentiating this equation with respect to s, and putting $s=1$ (with $S'(1)=\bar{n}$ and $S(1)=1$), we obtain

$$\bar{n} = \{ 1 - [(1+\epsilon)/2]^{V} \} / \{1 - 2[(1+\epsilon)/2]^{V} \}$$

$$= \{ 1 - [1 - e^{-TV}/2]^{V} \} / \{ 1 - 2 [1 - e^{-TV}/2]^{V} \}$$

Hence $\bar{n}$ is finite provided

$$T < - 1/v \ln (2 - 2 \times 2^{-\frac{1}{V}})$$

The upper bound equals the maximum sum-cutoff rate derived in section 7.3. The complexity $\bar{n}$ versus the total throughput of the users is plotted in figure 7.4.1.

Unfortunately, S(s) is unbounded for s>1, which will be proved in the remainder of this section.

Theorem 7.4.1

If S(s) is finite in the interval $(1,1+\epsilon)$ for some $\epsilon>0$, then S(s) is finite for all s>1.

Proof

Rearranging the terms in the previously derived equation for S(s), we have

$$S(s^2) = S(s) - \sum_{r=0}^{V-1} {}_{V}C_{r} \, \epsilon^{r}(1-\epsilon)^{V-r} \, s/(1-2^{-(V-r)} + 2^{-(V-r)} s) \times S((1-2^{-(V-r)} + 2^{-(V-r)} s)^2)$$

For all s>1, the argument s^2 of the function S on the left hand side can be readily shown to be larger than all arguments of the function S on the right hand side. The largest argument on the right hand side is $(1/2+1/2\ s)^2$. Assume that S(s) is finite up to

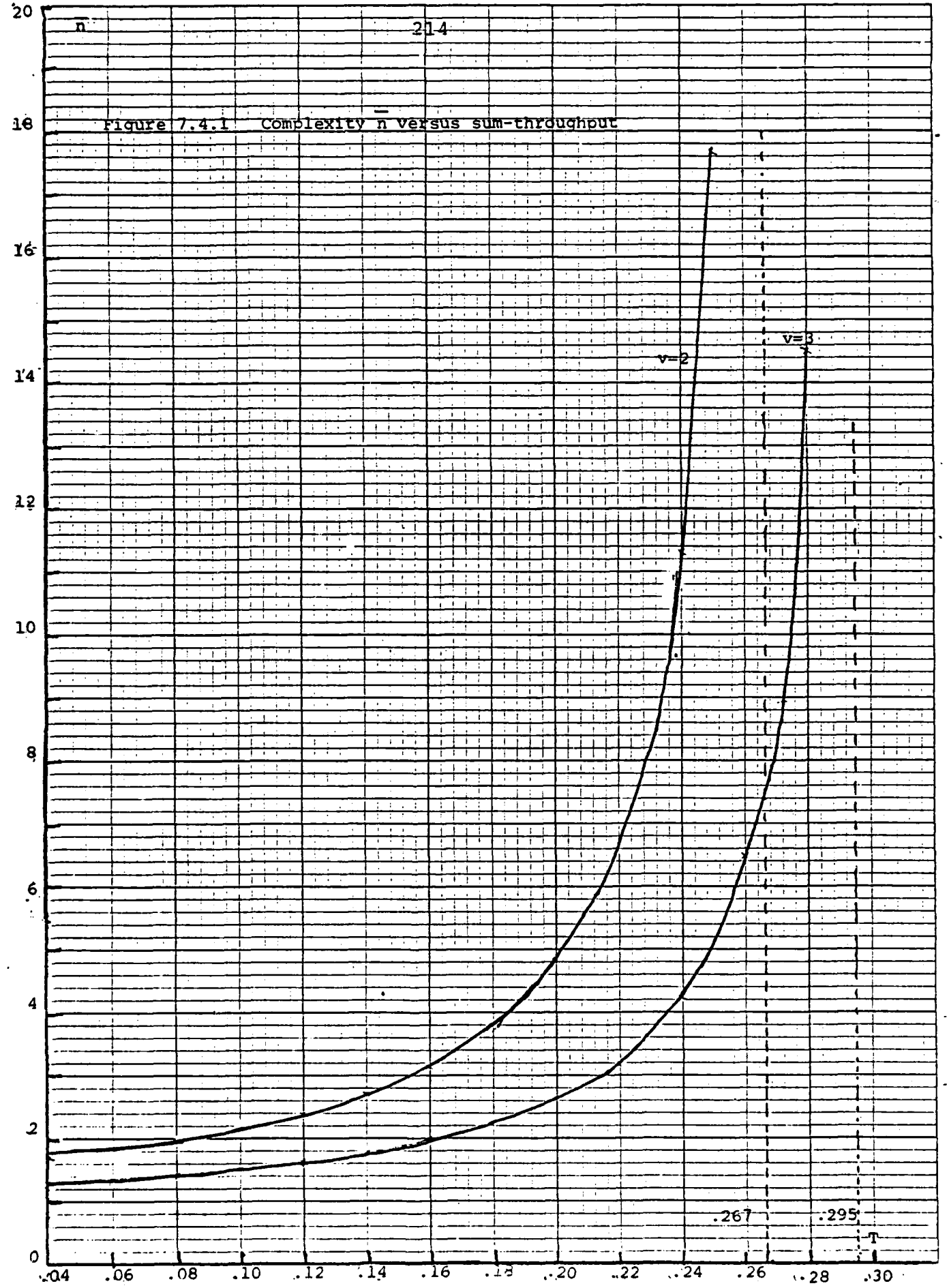


Figure 7.4.1 Complexity $\bar{n}$ versus sum-throughput

$s=s'$. Letting $s'>(1/2+1/2,s)^2$ (or $s<2\sqrt{s'}-1$, thus $s^2<4s'-4\sqrt{s'}+1$) implies that all terms on the right hand side are finite. Thus all values of $S(s)$ up to $s=4s'-4\sqrt{s'}+1$ are finite, since the left hand side is a sum of $v+1$ finite terms. Repeating the above argument, we have $S(s)$ finite for all finite s, provided that $S(s)$ is finite in the interval $(1,1+\epsilon)$ for some $\epsilon>0$.

Q.E.D.

Theorem 7.4.1

$S(s)$ does not exist for $s>1$.

Proof

Modify the branching process considered in this chapter by having a genie that always reveals the correct state (thus $n_j=1$) except when all v channel symbols are erased (thus $n_j=2n_{j-1}$), which occurs with probability ϵ^v.

For the process with a genie the steady state probability of the number of active states is

$$P(n) = (1-\epsilon^v)\ (\epsilon^v)^{\log_2 n} \qquad n = 2^k, \quad k \text{ integer}$$

$$= (1-\epsilon^v)\ n^{\log_2 \epsilon^v}$$

Its generating function does not exist for $s>1$. Obviously, the number of active states without genie is greater, and it will not have a generating function for $s>1$ either.

7.5 The collision channel with additive Gaussian noise

This section shows that the coding that is originally intended for correcting erasures is effective for combatting additive Gaussian channel noise. Therefore, signaling power can be reduced compared with the case of no coding at hardly any extra cost.

We shall assume that antipodal signaling is used. The results obtained in this section can be extended easily to other modulation techniques. The channel can be modeled as the memoryless binary input, Gaussian or erasure output channel shown in figure 7.5.1a. The capacity of this channel is

$C = (1 - \epsilon)$ of the capacity of the channel in figure 7.5.1b

$$= (1 - \epsilon)\left\{ -1/2 \log_2(2\pi e) - \int_{-\infty}^{\infty} P(y)\log_2 P(y)\, dy \right\}$$

where

$$P(y) = (P_1(y) + P_{-1}(y))/2$$

and

$$P_\ell(y) = \exp\left[-(y - \ell\,(2E_s/N_0)^{1/2})^2/2 \right] /\sqrt{2\pi}$$

The cut-off rate can be shown to be

$$R_0 = -\log_2\left[\epsilon + (1-\epsilon)(1+e^{-E_s/N_0})/2 \right] \qquad 7.5.1$$

Since sequential decoders are used for decoding, we concentrate on the study of the cut-off rate. Figure 7.5.2 gives a plot of R_0

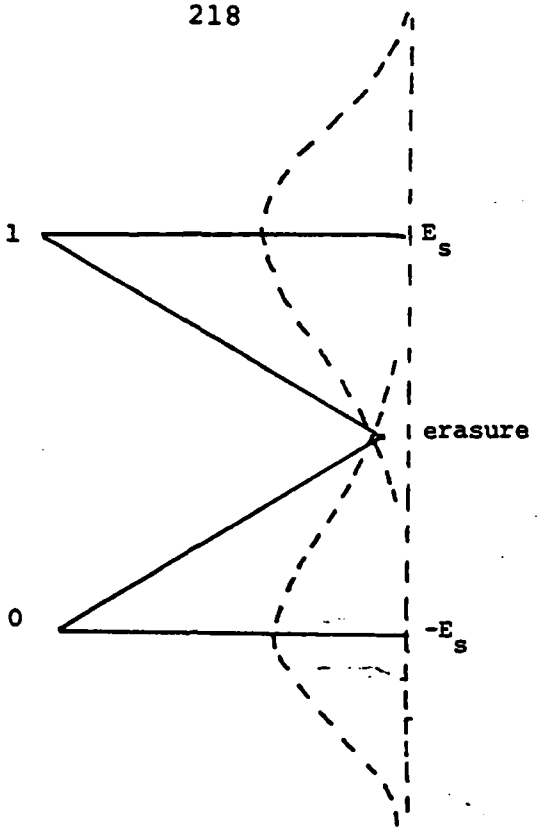


Figure 7.5.1a The binary input, erasure or Gaussian output channel

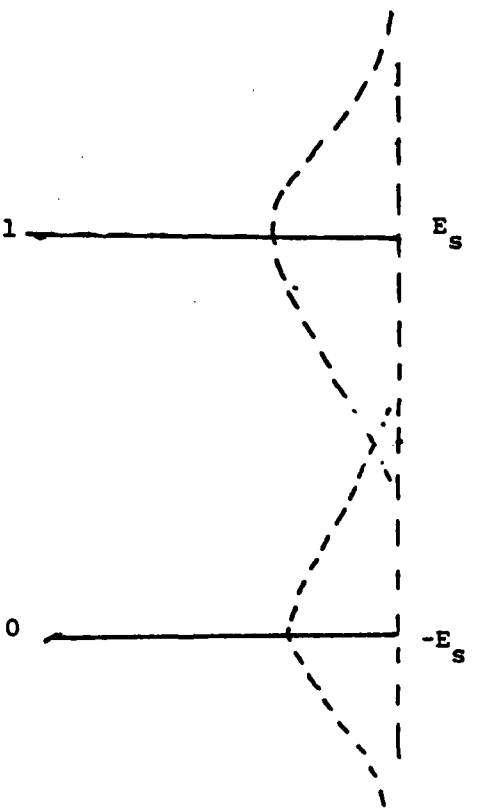


Figure 7.5.1b The binary input Gaussian output channel

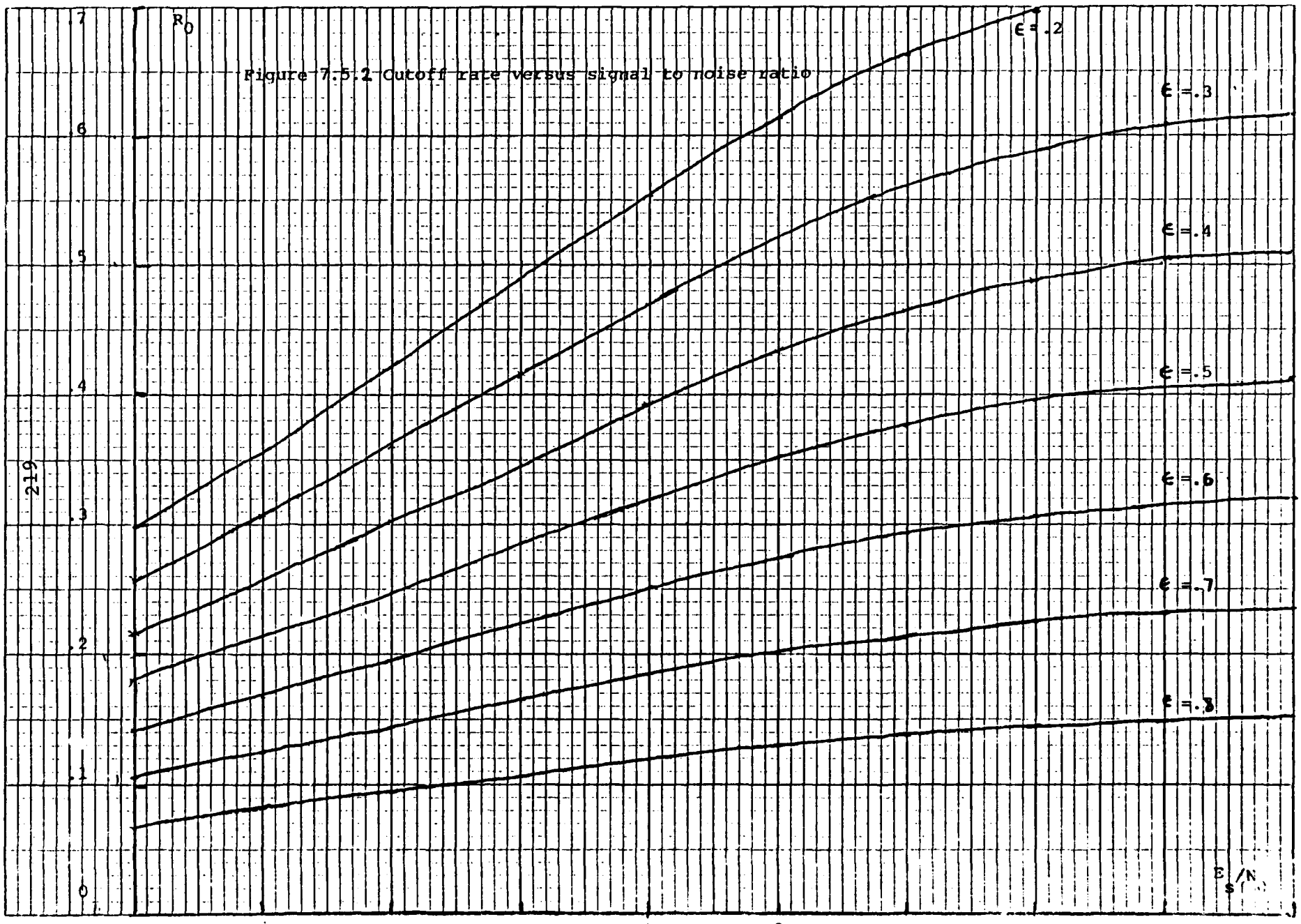


Figure 7.5.2 Cutoff rate versus signal to noise ratio

as a function of E_s/N_0 for several values of ϵ. The value of R_0 is seen to saturate at

$$R_0^* = -\log_2((1+\epsilon)/2)$$

for large values of E_s/N_0. This occurs when the occurrence of erasures dominates the effect of the Gaussian noise of the channel.

Suppose we allow Gaussian noise to decrease R_0 from R_0^* by a fraction f, such that

$$R_0 = f R_0^* = -f \log_2(1 - e^{-TV}/2) \qquad 7.5.2$$

Consequently, the maximum sum-throughput achievable by sequential decoding is

$$T = -1/v \ln(2 - 2 \times 2^{-\frac{1}{Rf}}) \qquad 7.5.3$$

This throughput as a function of f is shown in figure 7.5.3 for v=2,3.

What remains to be found is the value of the minimum E_s/N_0 that is required in order to achieve the maximum value of T for fixed v and f. Eliminating R_0 and ϵ in (7.5.1) by the substitutions

$$R_0 = -f \log_2(1 - e^{-TV}/2)$$

$$\epsilon = 1 - e^{-TV}$$

with T given in (7.5.3), we have after some simplification

Figure 7.5.3 Maximum throughput T for given f and v

f	v=2	v=3
0.40	--	.044
0.45	--	.073
0.50	--	.100
0.55	.034	.126
0.60	.065	.149
0.65	.095	.171
0.70	.124	.192
0.75	.150	.211
0.80	.176	.230
0.85	.200	.247
0.90	.224	.264
0.95	.246	.280
1.00	.267	.295

$$E_s/N_0 = -\ln\left[\left(2^{-1/v} - 2^{-1/vf}\right) / \left(1 - 2^{-1/vf}\right)\right]$$

which is shown in figure 7.5.4 for values of v and f. The figure shows that 4 or 5 dB of signal to noise ratio is sufficient for operation. This compares favorably with the typical 10.5 dB required for uncoded antipodal signaling for a bit error rate of 10^{-6}.

Figure 7.5.4 Power required to achieve maximum T for given f and v

f	E_s/N_o (dB) v=2	v=3
0.35	-2.02	-2.54
0.40	-1.52	-1.97
0.45	-1.03	-1.42
0.50	-0.55	-0.89
0.55	-0.07	-0.36
0.60	0.42	0.16
0.65	0.91	0.68
0.70	1.42	1.22
0.75	1.95	1.78
0.80	2.53	2.38
0.85	3.17	3.04
0.90	3.95	3.84
0.95	5.02	4.93
0.98	6.12	6.05

Appendix 7.1

This appendix derives the sum-throughput of the unslotted collision channel for which we may recover the front part of a packet up to the point where it starts to collide with another packet. Let 1-p be the probability that a user puts a packet in a slot. The probability that a packet is collision free is $p^{2(M-1)}$. The probability that the packet is totally lost is $1-p^{M-1}$. Let D_j be the difference in the starting time between user i and user j, as shown in figure 7.1.2. The probability that at least a fraction f of a packet of user i is recovered is

$$P_f = \prod_{\substack{j=1 \\ j \neq i}}^{M} [\ P(D_j \leq f)\ P(\text{no packet for user j in the two slots that overlaps with the packet of user i})$$

$$+ P(D_j > f)\ P(\text{no packet for user j in the first slot of the two slots that overlap with the packet of user i})\ ]$$

$$= [fp + (1-f)\ p]^{M-1}$$

Hence

$$P(f) = -d\ P_f/df = (M-1)\ (1-p)\ p^{M-1}\ (1-f(1-p))^{M-2}$$

Therefore, the average fraction of a packet that is recovered is

$$\int_0^1 P(f)\ f\ df + 1.p^{2(M-1)}$$

$$= (1-p^M)\ p^{M-1}\ /[(M-1)(1-p)]$$

The sum of the throughput for the M users is

$$T = M\,(1-p^{M})\,p^{M-1} / (M-1)$$

which has a maximum value of

$$M/(M-1)\quad M/(2M-1)\quad [\,(M-1)/(2M-1)\,]^{(M-1)/M}$$

which occurs at

$$p = [\,(M-1)/(2M-1)\,]^{1/M}$$

For large M, this throughput approaches 1/4.

Appendix 7.2

This appendix derives the throughput of the unslotted collision channel with super-packeting. Let q=1-p be the probability that each user puts a super-packet in a super-slot. For user i

P(a packet is collision free with user j)

= P(a super-slot for user j does not start within the duration of the packet).
P(no super-packet in the superslot of user j that overlaps with the packet)

+ P(a super-slot for user j starts within the duration of the packet)
P(no super-packet in either super-slots of user j that overlap with the packet)

$$= (u-1)/u\ p + 1/u\ p^2$$

Therefore, a packet is successfully transmitted with probability

$$\{ (u-1)/u\ p + 1/u\ p^2 \}^{M-1}$$

and the sum throughput is

$$T = M (1-p) \{ (u-1)/u\ p + 1/u\ p^2 \}^{M-1}$$

Letting q=f/M and assuming large value of M give the value T of

$$= f \{ (u-1)/u\,(1 - f/M) + 1/u\,(1 - f/M)^2 \}^{M-1}$$

$$= f \{ 1 - 1/M ((u-1)/u\, f + 2/u\, f) \}^{M-1}$$

$$= f \{ 1 - 1/M ((u+1)/u\, f) \}^{M-1}$$

$$= f\, e^{-f(u+1)/u}$$

which has a mximum value of $e^{-1}\, u/(u+1)$ when $f=u/(u+1)$. Thus the capacity approaches e^{-1} for large u, which is the same as the capacity for the slotted channel.

References

1. E. C. Van der Meulen, " A survey of multi-way channels in information theory, 1961-1975," IEEE Trans. on Inform. Theory, Vol IT-23, pp. 1-37, Jan. 1977.
2. N. Abramson, " The Aloha system - another alternative for computer communications," Fall Joint Computer Conference, AFIPS Conf. Proc., Vol 37, 1970.
3. J. L. Massey, " Collision-resolution algorithms and random-access communication," in Multi-User Communication Systems, G. Longo, Ed CISM courses and lecture series, No. 265, New York: Springer-Verlag, 1981, pp. 73-137.
4. A. R. Cohen, J. A. Heller, A. J. Viterbi, " A new coding technique for asynchronous multiple access communication," IEEE Trans. on Comm. Vol COM-19 pp. 849-855, Oct. 1971.
5. R. L. Pickholtz, D. L. Schilling, L. B. Milstein, " Theory of spread-spectrum communication - a tutorial," IEEE Trans. on Comm. Vol. COM-30, May 1982.
6. J. L. Massey, " The capacity of the collision channel without feedback," correspondence.
7. T. M. Cover, R. J. McEliece, E. C. Posner, " Asynchronous multiple-access channel capacity," IEEE Trans. on Inform. Theory Nov. 80 pp. 291-298.
8. L. Györfi, I. Kerekes, " A block code for noiseless asynchronous multiple-access OR channel," IEEE Trans. on Inform.Theory Nov. 81 pp. 788-791.

9. R. G. Gallager, Information Theory and Reliable Communication, Wiley, 1968.

10. A. J. Viterbi, J. K. Omura, Principles of Digital Communication and Coding, New York: McGraw Hill, 1979.

11. L. Kleinrock, Y. Yemini, " An optimal adaptive scheme for multiple access broadcast communication," ICC Conf. Proc. pp 7.2.1-7.2.5, 1978.

12. D. Blackwell, L. Breiman, A. J. Thomasian, " The capacity of a class of channels," Ann. Math. Stat. 30, 1229-1241.

13. I. G. Stiglitz, " Coding for a class of unknown channel," IEEE Trans. Inform. Theory, IT-12, 189-195.

14. R. J. McEliece, W. E. Stark, " An information theoretic study of communication in the presence of jamming," ICC Conf. Proc. pp. 45.3.1-45.3.6, 1981

15. S. Karlin, H. M. Taylor, A First Course in Stochastic Processes, Academic Press, London, 1975.

16. J. Wolfowitz, Coding Theorems of Information Theory, 2d ed., Springer-Verlag and Prentice-Hall, Englewood Cliffs, N.J., 1957.

17. J. P. Oderwalder, "Optimal decoding of convolutional codes", Ph.D. Dissertation, University of California, Los Angeles, 1970.

18. M. Avriel, Nonlinear Programming: Analysis and Methods, Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1976.

19. V. W. S. Chan, "A multiple-user random-access optical communication system," ICC conference records, ICC 79, pp. 1.4.1-1.4.5.

20. Capetanakis, J. I., "The multiple access broadcast channel: protocol and capacity considerations," Ph.D. Thesis, Dept. of Electrical Engineering and Computer Science, M.I.T. Cambridge, MA, August 1977.